



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 105371

TO: Rebecca Prouty
Location: CM1/10A13/10D01
Art Unit: 1652
Friday, October 10, 2003

Case Serial Number: 09/884889

From: Edward Hart
Location: Biotech-Chem Library
CM1-6B02
Phone: 305-9203

edward.hart@uspto.gov

Search Notes

Examiner Prouty,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

THIS PAGE BLANK (USPTO)

STIC-Biotech/ChemLib

105371

From: Prouty, Rebecca
Sent: Monday, October 06, 2003 8:48 AM
To: STIC-Biotech/ChemLib
Subject: Sequence Search

Art Unit 1652, 10A13
Mailbox: 10D01
308-4000
Serial Number: 09/884,889

Please search and interference search SEQ NOS: 5-8

RECEIVED
OCT - 6 2003
STIC-BIOTECH DIVISION
(STIC)

Edward Hart
Technical Info. Specialist
STIC/Biotech
DRI 6502 Tel: 305-9203

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 10/11/03
Date Completed: 10/11/03
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH: 2
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: 04/02
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)



STIC SEARCH RESULTS

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher* or contact:

Mary Hale, Information Branch Supervisor
308-4258, CM1-1E01

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library CM1 - Circ. Desk



THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2003, 07:56:28 ; Search time 127.677 Seconds
(without alignments)
7819.780 Million cell updates/sec

Title: US-09-884-889-5
Perfect score: 2262
Sequence: 1 atgaataacgcacgcgtga.....gttcgacgcgtgcgtgaa 2262

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2262	100.0	2262	2	US-08-674-887A-5
2	2262	100.0	2262	3	US-08-951-844-5
3	2262	100.0	2262	4	US-09-412-347-5
4	672.2	29.7	2238	2	US-08-674-887A-7
5	672.2	29.7	2238	3	US-08-951-844-7
6	672.2	29.7	2238	4	US-09-412-347-7
7	618.6	27.3	2331	1	US-08-418-782-20
8	618.6	27.3	2331	2	US-08-852-219-20
9	618.6	27.3	4403765	3	US-09-103-840A-2
10	618.6	27.3	4411529	3	US-09-103-840A-1
11	606.6	26.8	2235	1	US-08-418-782-1
12	606.6	26.8	2235	1	US-08-228-662-1
13	606.6	26.8	2235	2	US-08-852-219-1
14	578.2	25.6	2221	1	US-08-418-782-2
15	578.2	25.6	2221	1	US-08-228-662-2
16	578.2	25.6	2221	2	US-08-852-219-2
17	578.2	25.6	4795	2	US-08-313-185-45
18	578.2	25.6	4795	3	US-09-082-614A-45
19	577.6	25.5	4794	2	US-08-459-499-8
20	514.8	22.8	2181	4	US-09-328-352-3989
21	137.2	6.1	620	2	US-08-757-653-144
22	137.2	6.1	620	2	US-08-757-653-146
23	137.2	6.1	620	2	US-08-757-653-148
24	137.2	6.1	620	2	US-08-757-653-150
25	137.2	6.1	620	4	US-08-520-946-144
26	137.2	6.1	620	4	US-08-520-946-146
27	137.2	6.1	620	4	US-08-520-946-148

c	28	137.2	6.1	620	4	US-08-520-946-150	Sequence 150, App
	29	135.6	6.0	620	2	US-08-757-653-143	Sequence 143, App
	30	135.6	6.0	620	2	US-08-757-653-145	Sequence 145, App
c	31	135.6	6.0	620	2	US-08-757-653-147	Sequence 147, App
c	32	135.6	6.0	620	2	US-08-757-653-149	Sequence 149, App
	33	135.6	6.0	620	4	US-08-520-946-143	Sequence 143, App
	34	135.6	6.0	620	4	US-08-520-946-145	Sequence 145, App
c	35	135.6	6.0	620	4	US-08-520-946-147	Sequence 147, App
c	36	135.6	6.0	620	4	US-08-520-946-149	Sequence 149, App
	37	91	4.0	391	3	US-09-034-205-4	Sequence 4, Appli
	38	91	4.0	391	3	US-08-934-097A-4	Sequence 4, Appli
	39	91	4.0	391	3	US-08-851-588-4	Sequence 4, Appli
	40	91	4.0	391	4	US-09-677-218B-4	Sequence 4, Appli
	41	91	4.0	391	3	US-09-677-192-4	Sequence 4, Appli
	42	89.4	4.0	391	3	US-09-034-205-1	Sequence 1, Appli
	43	89.4	4.0	391	3	US-09-034-205-2	Sequence 2, Appli
	44	89.4	4.0	391	3	US-08-934-097A-1	Sequence 1, Appli
	45	89.4	4.0	391	3	US-08-934-097A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-674-887A-5
; Sequence 5, Application US/08674887A
; Patent No. 5939300
; GENERAL INFORMATION:
; APPLICANT: Robertson, Dan E.
; APPLICANT: Sanyal, Indrajit
; TITLE OF INVENTION: CATALASES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P. C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/674,887A
; FILING DATE: 03-JUL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09015/002001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2262 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...2259
; NAME: US-08-674-887A-5

Query Match 100.0%; Score 2262; DB 2; Length 2262;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGAATACGCATCCGCTGACGATCTACACAGTAGCTTGCAGCAAAAGATGCAGAGCAATTT	60
Db	1	ATGAATACGCATCCGCTGACGATCTACACAGTAGCTTGCAGCAAAAGATGCAGAGCAATTT	60
QY	61	GTTCCCTTGGTATGCCAAGGCATAGAGCAATAAGGGAGAGAGCTATGACGGTAAATGT	120
Db	61	GTTCCCTTGGTATGCCAAGGCATAGAGCAATAAGGGAGAGAGCTATGACGGTAAATGT	120
QY	121	CCTGTCATGACCGTGGTAAACACTGCACCGGTACTTCCAACAAGAGATTGTTGGCCGGAA	180
Db	121	CCTGTCATGACCGTGGTAAACACTGCACCGGTACTTCCAACAAGAGATTGTTGGCCGGAA	180
QY	181	GGGTTGAACCTGGATATTTTGCATCAGCAAGATCGCAATCAGACCCGATGGATCCGGAT	240
Db	181	GGGTTGAACCTGGATATTTTGCATCAGCAAGATCGCAATCAGACCCGATGGATCCGGAT	240
QY	241	TTCAACTACCGTGAAGAAGTACGCAAGCTCGATTTTCGACGCGTGAAGAAGATGTCAC	300
Db	241	TTCAACTACCGTGAAGAAGTACGCAAGCTCGATTTTCGACGCGTGAAGAAGATGTCAC	300
QY	301	CGGTTGATGACCGATACCAAGAGTGTGGCCCGCTGACTGGGGCACTACGGCGGTTG	360
Db	301	CGGTTGATGACCGATACCAAGAGTGTGGCCCGCTGACTGGGGCACTACGGCGGTTG	360
QY	361	ATGATCCGTATGGCTTGGCACTCCGCTGGCACTACCGTATTCGCTGATGCCCTGGGGC	420
Db	361	ATGATCCGTATGGCTTGGCACTCCGCTGGCACTACCGTATTCGCTGATGCCCTGGGGC	420
QY	421	GGTGGTACCGGAAGCCAGCGCTTTGCAACCGCTCAACTCCTGGCGGACAACGTCAGGCTG	480
Db	421	GGTGGTACCGGAAGCCAGCGCTTTGCAACCGCTCAACTCCTGGCGGACAACGTCAGGCTG	480
QY	481	GATAAAGCGCGCTCTGCTGTGGCCCATCAAGAAGAAAGTACGSCAACAAAATCAGCTGG	540
Db	481	GATAAAGCGCGCTCTGCTGTGGCCCATCAAGAAGAAAGTACGSCAACAAAATCAGCTGG	540
QY	541	GCAGACCTGATGATCTTGGCTGGCACCGTGGCTTATGAGTCCATGGGCTTACCTGCTTAC	600
Db	541	GCAGACCTGATGATCTTGGCTGGCACCGTGGCTTATGAGTCCATGGGCTTACCTGCTTAC	600
QY	601	GGCTTCTCTTTCGGCCGCGTGATATTGTTGGAAACCCGAAAAAGATATCTACTGGGCTGAC	660
Db	601	GGCTTCTCTTTCGGCCGCGTGATATTGTTGGAAACCCGAAAAAGATATCTACTGGGCTGAC	660
QY	661	GAAAAACAGTGGCTGGCACCTTCTGACGACGCTACGGCAGCTGACACAGCCAGACACC	720
Db	661	GAAAAACAGTGGCTGGCACCTTCTGACGACGCTACGGCAGCTGACACAGCCAGACACC	720
QY	721	ATGAAAAACCCGCTGGCGGCTGTCCAAATGGGCTGATCTATGTGAAACCCGGAAGGTGT	780
Db	721	ATGAAAAACCCGCTGGCGGCTGTCCAAATGGGCTGATCTATGTGAAACCCGGAAGGTGT	780
QY	781	AACGGCCACCTGATCCGCTGAGAACCGCACAGAGTACTTGAACCTTTCGCCCGGTATG	840
Db	781	AACGGCCACCTGATCCGCTGAGAACCGCACAGAGTACTTGAACCTTTCGCCCGGTATG	840
QY	841	CGGATGAACACGAAAAAACCGCAGCCCTCACAGCTGGCGGCCACACCCGTCGGTAATGT	900
Db	841	CGGATGAACACGAAAAAACCGCAGCCCTCACAGCTGGCGGCCACACCCGTCGGTAATGT	900
QY	901	CACGGTAAATGGCAATGCCTCTGGTTAGCCCTGACCACAAAGCCTCTGACGTTGAAAC	960
Db	901	CACGGTAAATGGCAATGCCTCTGGTTAGCCCTGACCACAAAGCCTCTGACGTTGAAAC	960
QY	961	CAGGGCTTAGTTGGGCAACCCCAACATGCAGGCAAGGCAAGCAACGCGCTGACCTCG	1020
Db	961	CAGGGCTTAGTTGGGCAACCCCAACATGCAGGCAAGGCAAGCAACGCGCTGACCTCG	1020
QY	1021	GGTATCAAGAGTGTGGACCAACCAACCCACGAAATTCGATATGGGCTATTTTCGACCTG	1080
Db	1021	GGTATCAAGAGTGTGGACCAACCAACCCACGAAATTCGATATGGGCTATTTTCGACCTG	1080

Db 2161 GAAAGTGACGCCAGGAGATAACGGCAGAGAGTTTCGTACAGACTTCGTCCGCCCTGG 2220
QY 2221 ACCAAGTGATGAACGCCGACCGTTTCGACGTCGCGTGTAA 2262
Db 2221 ACCAAGTGATGAACGCCGACCGTTTCGACGTCGCGTGTAA 2262

RESULT 2
US-08-951-844-5
; Sequence 5, Application US/08951844
; Patent No. 6074860
; GENERAL INFORMATION:
; APPLICANT: Robertson et al.
; TITLE OF INVENTION: Catalases
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSER: CARELLA, BYRNE, HAIN, GILFILLAN,
; ADDRESSER: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/951,844
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/674,887
; FILING DATE: July 3, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles J. Herron
; REGISTRATION NUMBER: 28,019
; REFERENCE/DOCKET NUMBER: 331400-55
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2262 NUCLEOTIDES
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: Oligonucleotide
US-08-951-844-5

Query Match 100.0%; Score 2262; DB 3; Length 2262;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAATAACGCATCCGCTGACGATCTACACAGTCTTCAGCAAGATGCGAGAGCATTT 60
Db 1 ATGAATAACGCATCCGCTGACGATCTACACAGTCTTCAGCAAGATGCGAGAGCATTT 60

QY 61 GTTCCCTTGGTATCGCCAGGATAGCAATTAAGGAGAGAGCTATGACCGGTAATGT 120
Db 61 GTTCCCTTGGTATCGCCAGGATAGCAATTAAGGAGAGAGCTATGACCGGTAATGT 120

QY 121 CTGTCTATCAGCGGTGAACACCTCGACCGTACTTCCAAACAAGATTGGTGGCCGAA 180
Db 121 CTGTCTATCAGCGGTGAACACCTCGACCGTACTTCCAAACAAGATTGGTGGCCGAA 180

QY 181 GGGTTGAACCTGGATATTTTGGATCAGCAAGATCGCAAAATCAGACCCGATGATCGGGAT 240
Db 181 GGGTTGAACCTGGATATTTTGGATCAGCAAGATCGCAAAATCAGACCCGATGATCGGGAT 240

QY 241 TTCAACTACCGTGAAGAAGTACGCAAGCTCGATTTTCGACCGCGCTGAAGAAGATGTCCAC 300

Db 241 TTCAACTACCGTGAAGAAGTACGCAAGCTCGATTTTCGACCGCTGAAGAAGATGTCCAC 300
QY 301 GCGTTGATGACCGATAGCCCAAGAGTGGTGGCCGCTGACTGGGGGCACTACGCGGTTTG 360
Db 301 GCGTTGATGACCGATAGCCCAAGAGTGGTGGCCGCTGACTGGGGGCACTACGCGGTTTG 360

QY 361 ATGATCCGATGGCTTGGCACTCCGCTGGCACTACCGTATTTGCTGATGGCGTGGGGC 420
Db 361 ATGATCCGATGGCTTGGCACTCCGCTGGCACTACCGTATTTGCTGATGGCGTGGGGC 420

QY 421 GGTGGTACCGGAAGCCAGCGCTTGGACCGCTCAACTCTCTGGCCGACAACGTCAGCCTG 480
Db 421 GGTGGTACCGGAAGCCAGCGCTTGGACCGCTCAACTCTCTGGCCGACAACGTCAGCCTG 480

QY 481 GATAAAGCGCGCTCTGCTGTGGCGGATCAAGAAGAAGTAGGGCAACAAAATCAGCTGG 540
Db 481 GATAAAGCGCGCTCTGCTGTGGCGGATCAAGAAGAAGTAGGGCAACAAAATCAGCTGG 540

QY 541 GCAGACCTGATGATTTGCTGGCACTCCGCTGGCTTTATGATGCTTACCTGCTTAC 600
Db 541 GCAGACCTGATGATTTGCTGGCACTCCGCTGGCTTTATGATGCTTACCTGCTTAC 600

QY 601 GCGTTCTCTTTGCGCGCGCTCGATATTTGGGAACCCGAAAAGATATCTACTGGGGTGAC 660
Db 601 GCGTTCTCTTTGCGCGCGCTCGATATTTGGGAACCCGAAAAGATATCTACTGGGGTGAC 660

QY 661 GAAAAAGAGTGGCTGGCACTTCTGACGAACGCTAGCGGACGTGAACAGCCAGAGACC 720
Db 661 GAAAAAGAGTGGCTGGCACTTCTGACGAACGCTAGCGGACGTGAACAGCCAGAGACC 720

QY 721 ATGGAAAAACCCGCTGGCGCTGTCCAAATGGGTCTGATCTATGTGAACCCGGAAGTGT 780
Db 721 ATGGAAAAACCCGCTGGCGCTGTCCAAATGGGTCTGATCTATGTGAACCCGGAAGTGT 780

QY 781 AACGGCCACCTGATCCGCTGAGAACCCGACAGCAGGTACTTGAACCTTCGCCCGCTATG 840
Db 781 AACGGCCACCTGATCCGCTGAGAACCCGACAGCAGGTACTTGAACCTTCGCCCGCTATG 840

QY 841 GCGATGAACGAGAAAAAACCCGACGCCCTCACAGCTGGCGGCCACACCGTCGTAATTTGT 900
Db 841 GCGATGAACGAGAAAAAACCCGACGCCCTCACAGCTGGCGGCCACACCGTCGTAATTTGT 900

QY 901 CACGGTAAATGGCAATGCCTCTGCGTTAGCCCTGACCCAAAGACCTCTGACGTTGAAAC 960
Db 901 CACGGTAAATGGCAATGCCTCTGCGTTAGCCCTGACCCAAAGACCTCTGACGTTGAAAC 960

QY 961 CAGGGCTTAGGTTGGGGCAACCCCAACATGCAGGGCAAGCAAGCAACGCCGTGACCTCG 1020
Db 961 CAGGGCTTAGGTTGGGGCAACCCCAACATGCAGGGCAAGCAAGCAACGCCGTGACCTCG 1020

QY 1021 GGTATCGAAGTGCTTGGACCAACCCCAACATTCGATATGGCTATTTTCGACCTG 1080
Db 1021 GGTATCGAAGTGCTTGGACCAACCCCAACATTCGATATGGCTATTTTCGACCTG 1080

QY 1081 CTGTTCCGGCTACAAATTTGGGAACCTGAAAAAGAGTCTCTGCCGTCGCCACCATTTGGAACCG 1140
Db 1081 CTGTTCCGGCTACAAATTTGGGAACCTGAAAAAGAGTCTCTGCCGTCGCCACCATTTGGAACCG 1140

QY 1141 ATTGACATCAAAAAGAAAAACAAGCCGGTTGACGCCAGCCGCTCTATTTCGCCCAAC 1200
Db 1141 ATTGACATCAAAAAGAAAAACAAGCCGGTTGACGCCAGCCGCTCTATTTCGCCCAAC 1200

QY 1201 CCGATCATGACCGATCGGATATGGGATAAAGGTAATCCGACCTATTCGCTATCTGC 1260
Db 1201 CCGATCATGACCGATCGGATATGGGATAAAGGTAATCCGACCTATTCGCTATCTGC 1260

QY 1261 GAAAAATTCATGCGCGATCTCTGAGTACTTCAAGAAAACTTTCGCGAGGCGTGGTCAAG 1320
Db 1261 GAAAAATTCATGCGCGATCTCTGAGTACTTCAAGAAAACTTTCGCGAGGCGTGGTCAAG 1320

QY 1321 CTGACGACCGCTGACCTGGCGCCGAAATCAGTTTACATCGCCCGGAAAGTTCGCGGCAAG 1380
Db 1321 CTGACGACCGCTGACCTGGCGCCGAAATCAGTTTACATCGCCCGGAAAGTTCGCGGCAAG 1380

```
Db 1371 CTGACGCACCGTGACCTGGGCCCCGAAATACGTTACATCGGCCCGGAAGTCCCGCAGAA 1380
Qy 1381 GACCTGATTTGGCAAGACCCGATTCGGSCAGSTAAACACCCGACTACTGCGAAGAGTGGTC 1440
Db 1381 GACCTGATTTGGCAAGACCCGATTCGGSCAGSTAAACACCCGACTACTGCGAAGAGTGGTC 1440
Qy 1441 AAGCAGAAATTCACAAAGTGGCTGAGCATTAGTGTGATGCTCCACCGCTTGGGAC 1500
Db 1441 AAGCAGAAATTCACAAAGTGGCTGAGCATTAGTGTGATGCTCCACCGCTTGGGAC 1500
Qy 1501 AGTCCCGTACTTATCGCGTTCCGATATGCGCGGCTGCTAACGCTGCGCGATTCGC 1560
Db 1501 AGTCCCGTACTTATCGCGTTCCGATATGCGCGGCTGCTAACGCTGCGCGATTCGC 1560
Qy 1561 TTGCCCCACAAAGAGTGGCAGCGCAACGAGCGGAGCGCTTGGGAAAGTGTGCTGAGC 1620
Db 1561 TTGCCCCACAAAGAGTGGCAGCGCAACGAGCGGAGCGCTTGGGAAAGTGTGCTGAGC 1620
Qy 1621 GTCACAGCAGATCTCTGCGCACCGCGGCTAGCATCGCGGACGTGATGCTTCTGGCC 1680
Db 1621 GTCACAGCAGATCTCTGCGCACCGCGGCTAGCATCGCGGACGTGATGCTTCTGGCC 1680
Qy 1681 GGTAGCGTAGGCATCGAAGACCGCGAAGCAGCAGGTACGATGTCGCGCTTCCCTTC 1740
Db 1681 GGTAGCGTAGGCATCGAAGACCGCGAAGCAGCAGGTACGATGTCGCGCTTCCCTTC 1740
Qy 1741 CTGAAAGCCCTGGCGATGCGACCGCGGAGATGACGACGAGTCTCTTCGACCGCTG 1800
Db 1741 CTGAAAGCCCTGGCGATGCGACCGCGGAGATGACGACGAGTCTCTTCGACCGCTG 1800
Qy 1801 GAGCCGCTGGCGATGCGCACTGCGCACTGGCAGAAAGAGTATGTTGAAGCCGGA 1860
Db 1801 GAGCCGCTGGCGATGCGCACTGCGCACTGGCAGAAAGAGTATGTTGAAGCCGGA 1860
Qy 1861 GAGATGCTGTGAGTGTGCGCAGCTGATGGCTTAACCGCGCGGAAATGACCGCTG 1920
Db 1861 GAGATGCTGTGAGTGTGCGCAGCTGATGGCTTAACCGCGCGGAAATGACCGCTG 1920
Qy 1921 CTGGCGGCTATGCGCGTACTGGGCACCAACTATGCTGGCAGCAACACGCGGTATTCAC 1980
Db 1921 CTGGCGGCTATGCGCGTACTGGGCACCAACTATGCTGGCAGCAACACGCGGTATTCAC 1980
Qy 1981 GATTGTGAAGCCAGTGTACCAAGCAGCTTTTGTGAACTGACCGATATGGGGAACAGC 2040
Db 1981 GATTGTGAAGCCAGTGTACCAAGCAGCTTTTGTGAACTGACCGATATGGGGAACAGC 2040
Qy 2041 TGGAAAGCCGTAGTGTAGCAACCGCTACGAAATCGCGACGCAAGACCGGTGCGTGAAG 2100
Db 2041 TGGAAAGCCGTAGTGTAGCAACCGCTACGAAATCGCGACGCAAGACCGGTGCGTGAAG 2100
Qy 2101 TGGACCGCTCGCGGCTGGATCTGGTATTTGGTTCCAACTGCGTACTGCGCTTTACGCA 2160
Db 2101 TGGACCGCTCGCGGCTGGATCTGGTATTTGGTTCCAACTGCGTACTGCGCTTTACGCA 2160
Qy 2161 GAAGTGTACGCCAGGAGCATAGCGGCGAGAGTTCGTAGAGACTTCGTGCGCGCGCTGG 2220
Db 2161 GAAGTGTACGCCAGGAGCATAGCGGCGAGAGTTCGTAGAGACTTCGTGCGCGCGCTGG 2220
Qy 2221 ACCAAAGTATGAACCGCGACCGTTTCGACGTGCGCTGTA 2262
Db 2221 ACCAAAGTATGAACCGCGACCGTTTCGACGTGCGCTGTA 2262
```

```
RESULT 3
US-09-412-347-5
: Sequence 5, Application US/09412347
: Patent No. 6410290
: GENERAL INFORMATION:
: APPLICANT: Robertson, Dan E.
: APPLICANT: Sanyal, Indrajit
: APPLICANT: Adhikari, Robert S.
: TITLE OF INVENTION: CATALASES
: NUMBER OF SEQUENCES: 8
```

```
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 4225 Executive Square, Suite 1400
: CITY: La Jolla
: STATE: CA
: COUNTRY: US
: ZIP: 92037
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows95
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/412,347
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/674,887
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Hallie, Ph.D., Lisa A.
: REGISTRATION NUMBER: 38,347
: REFERENCE/DOCKET NUMBER: 09015/002001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 619/678-5070
: TELEFAX: 619/678-5099
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2262 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: Coding Sequence
: LOCATION: 1...2259
: US-09-412-347-5
```

```
Query Match 100.0%; Score 2262; DB 4; Length 2262;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAATAACGCATCCGTGAGATCTACACAGTAGTGTGACGAAAGATGCGAGCATTT 60
Db 1 ATGAATAACGCATCCGTGAGATCTACACAGTAGTGTGACGAAAGATGCGAGCATTT 60
Qy 61 GTTCCCTTGGTATCGCCAAAGCATAGAGCAATAAGGGAGAGAGCTATGAGCGGTAAATGT 120
Db 61 GTTCCCTTGGTATCGCCAAAGCATAGAGCAATAAGGGAGAGAGCTATGAGCGGTAAATGT 120
Qy 121 CCTGTCATGACACGGTGTAAACACCTCGACCGGTACTTCCAAACAAAGATTGGTGGCCGGA 180
Db 121 CCTGTCATGACACGGTGTAAACACCTCGACCGGTACTTCCAAACAAAGATTGGTGGCCGGA 180
Qy 181 GGGTTGAACCTGGATATTTTCATCAGCAAGATCGCAATACAGACCGATGGATCCGGAT 240
Db 181 GGGTTGAACCTGGATATTTTCATCAGCAAGATCGCAATACAGACCGATGGATCCGGAT 240
Qy 241 TTCAACTACCGTGAAGAAGTACGCAAGCTCGATTTCCACGCGCTGAAGAAGATGTCCAC 300
Db 241 TTCAACTACCGTGAAGAAGTACGCAAGCTCGATTTCCACGCGCTGAAGAAGATGTCCAC 300
Qy 301 CGGTTGATGACCGATAGCCAAAGAGTGGTGGCCCGCTGACTGGGGGCACTACGGCGGTTTG 360
Db 301 CGGTTGATGACCGATAGCCAAAGAGTGGTGGCCCGCTGACTGGGGGCACTACGGCGGTTTG 360
Qy 361 ATGATCCGTATGGCTTGGCACTCCGCTGGCACCTTACCGTATTTGCTGATGCCGCTGGGGC 420
Db 361 ATGATCCGTATGGCTTGGCACTCCGCTGGCACCTTACCGTATTTGCTGATGGCGCTGGGGC 420
Qy 421 GGTGGTACCGGAAGCCAGCGCTTTGGCACCGCTCAACTCCCTGGCGGGAACAACGTCAGCTG 480
Db 421 GGTGGTACCGGAAGCCAGCGCTTTGGCACCGCTCAACTCCCTGGCGGGAACAACGTCAGCTG 480
```

Qy	481	GATAAAGCGCGCGTCTGCTGTGGCGGATCAAGAAGAGTACGGCAACAAAATCAGCTGG	540
Db	481	GATAAAGCGCGCGTCTGCTGTGGCGGATCAAGAAGAGTACGGCAACAAAATCAGCTGG	540
Qy	541	GCAGACTGATGATTCCTGGCTGCGACCGTGGCTTATCAGTCCATGGCTTACCTGGCTTAC	600
Db	541	GCAGACTGATGATTCCTGGCTGCGACCGTGGCTTATGAGTCCATGGCTTACCTGGCTTAC	600
Qy	601	GGCTTCTCTTTCGGCCGCGTGCATATTTTGGGAACCGAAAAAGATATCTACTGGGTGAC	660
Db	601	GGCTTCTCTTTCGGCCGCGTGCATATTTGGGAACCGAAAAAGATATCTACTGGGTGAC	660
Qy	661	GAAGAAGTGGCTGGACCTCTTGACGAACCGTACGGGACGTGAACAGCCAGAGACC	720
Db	661	GAAGAAGTGGCTGGACCTCTTGACGAACCGTACGGGACGTGAACAGCCAGAGACC	720
Qy	721	ATGGAAAACCCGCTGGCGGCTGTCCAAATGGGTCTGATCTATCTGAACCCGGAAGTGTT	780
Db	721	ATGGAAAACCCGCTGGCGGCTGTCCAAATGGGTCTGATCTATCTGAACCCGGAAGTGTT	780
Qy	781	AACGGCCACCCTGATCCGCTGAGAACCGCACAGCAGTACTTTGAAACCTTTCGCCCGTATG	840
Db	781	AACGGCCACCCTGATCCGCTGAGAACCGCACAGCAGTACTTTGAACCTTTCGCCCGTATG	840
Qy	841	GCAGTGAACGACGAANAACCGCAGCCCTCACAGCTGGGGCCACACCGTCGGTAATGT	900
Db	841	GCAGTGAACGACGAANAACCGCAGCCCTCACAGCTGGGGCCACACCGTCGGTAATGT	900
Qy	901	CACGGTAATGGCAATGCCCTGCGTTAGCCCCCTGACCCAAAAGCCCTGACGTTGAAAC	960
Db	901	CACGGTAATGGCAATGCCCTGCGTTAGCCCCCTGACCCAAAAGCCCTGACGTTGAAAC	960
Qy	961	CAGGGCTTAGGTTGGGCAACCCCAACATCGAGGGCAAGGCAAGCAACGCCGTGACCTCG	1020
Db	961	CAGGGCTTAGGTTGGGCAACCCCAACATCGAGGGCAAGGCAAGCAACGCCGTGACCTCG	1020
Qy	1021	GGTATCGAAGGTGCTTGGACACCAACCCCAAGAAATTCGANTTGGGTATTTTCGACCTG	1080
Db	1021	GGTATCGAAGGTGCTTGGACACCAACCCCAAGAAATTCGANTTGGGTATTTTCGACCTG	1080
Qy	1081	CTGTTCGGCTACAAATGGGAACGTGAAAACAGTCTTCGCCGTGCCACCAATTCGGAAACG	1140
Db	1081	CTGTTCGGCTACAAATGGGAACGTGAAAACAGTCTTCGCCGTGCCACCAATTCGGAAACG	1140
Qy	1141	ATTGACATCAAAAGGAAAAACAAGCCGGTTGAGCGCCAGCGACCCCTCTATTTCGCCACAAC	1200
Db	1141	ATTGACATCAAAAGGAAAAACAAGCCGGTTGAGCGCCAGCGACCCCTCTATTTCGCCACAAC	1200
Qy	1201	CCGATCATGACGATGGGGATATGGGATTAAGGTAAATCCGACCTATCCGGTATCTGTC	1260
Db	1201	CCGATCATGACGATGGGGATATGGGATTAAGGTAAATCCGACCTATCCGGTATCTGTC	1260
Qy	1261	GAAAAATTATGCCGATCTGAGTACTTCAAGAAACCTTTCGGAGGCGTGGTTCAG	1320
Db	1261	GAAAAATTATGCCGATCTGAGTACTTCAAGAAACCTTTCGGAGGCGTGGTTCAG	1320
Qy	1321	CTGACGACCGGTGACCTGGGCCCGGAATACGTTTACATCGCCCGGAAGTTCGCGGAGAA	1380
Db	1321	CTGACGACCGGTGACCTGGGCCCGGAATACGTTTACATCGCCCGGAAGTTCGCGGAGAA	1380
Qy	1381	GACCTGATTTGGCAAGACCCGATTCGCGCAGGTAAACACCGACTACTTCGGAAGATGGTC	1440
Db	1381	GACCTGATTTGGCAAGACCCGATTCGCGCAGGTAAACACCGACTACTTCGGAAGATGGTC	1440
Qy	1441	ARGCAGAAAATTCACAAAAGTGGCCTGAGCATTTAGTGAGATGGTCTCCACCGCTTGGGAC	1500
Db	1441	ARGCAGAAAATTCACAAAAGTGGCCTGAGCATTTAGTGAGATGGTCTCCACCGCTTGGGAC	1500
Qy	1501	AGTGCCCGTACTTATCCGGTTCCGATATCGCGGGTCTTAAGCGTCCCGCATTCGC	1560
Db	1501	AGTGCCCGTACTTATCCGGTTCCGATATCGCGGGTCTTAAGCGTCCCGCATTCGC	1560
Qy	1561	TTGGCCCCCAGAACGAGTGGCAGGCAACGAGCCGAGGCGCTGGCGAAAGTCTGAGC	1620

[illegible]

```

1  RESULT 4
2  US-08-674-887A-7
3  : Sequence 7, Application US/08674887A
4  : Patent No. 5939300
5  : GENERAL INFORMATION:
6  : APPLICANT: Robertson, Dan E.
7  : APPLICANT: Sanyal, Indrajit
8  : APPLICANT: Adhikari, Robert S.
9  : TITLE OF INVENTION: CATALASES
10 : NUMBER OF SEQUENCES: 8
11 : CORRESPONDENCE ADDRESS:
12 : ADDRESS: Fish & Richardson P.C.
13 : STREET: 4225 Executive Square, Suite 1400
14 : CITY: La Jolla
15 : STATE: CA
16 : COUNTRY: US
17 : ZIP: 92037
18 : COMPUTER READABLE FORM:
19 : MEDIUM TYPE: Diskette
20 : COMPUTER: IBM Compatible
21 : OPERATING SYSTEM: Windows95
22 : SOFTWARE: FASTSEQ for Windows Version 2.0
23 : CURRENT APPLICATION DATA:
24 : APPLICATION NUMBER: US/08/674,887A
25 : FILING DATE: 03-JUL-1996

```

PRIOR APPLICATION DATA:									
APPLICATION NUMBER:									
FILING DATE:									
ATTORNEY/AGENT INFORMATION:									
NAME: Halle, Ph.D., Lisa A.									
REGISTRATION NUMBER: 38,347									
REFERENCE/DOCKET NUMBER: 09015/002001									
TELECOMMUNICATION INFORMATION:									
TELEPHONE: 619/678-5070									
TELEFAX: 619/678-5099									
INFORMATION FOR SEQ ID NO: 7:									
SEQUENCE CHARACTERISTICS:									
LENGTH: 2238 base pairs									
TYPE: nucleic acid									
STRANDEDNESS: double									
TOPOLOGY: linear									
MOLECULE TYPE: cDNA									
FEATURE:									
NAME/KEY: Coding Sequence									
LOCATION: 1...2235									
US-08-674-887A-7									
Query Match 29.7%; Score 672.2; DB 2; Length 2238;									
Best Local Similarity 59.5%; Pred. No. 3.7e-179;									
Matches 1277; Conservative 0; Mismatches 808; Indels 60; Gaps 6;									
Qy	150	CGGTACTTCCAAACAAGATTGGTGGCGGAAGGTTGAACCTGGATATTTTCATCAGCA	209						
Db	99	CGGCACCAAAACAGGATTTGGTGGCCCAACATGCTCAACCTGGCATCTTACGCCAACA	158						
Qy	210	AGATCGCAATACAGCCCGATGGATCGCGGATTTCAACTACCGTGAAGAAGTACCGAAGCT	269						
Db	159	TTCATCGTATCGGACCCCAACGACCGGATTTTGACTATGCCGAAGAGTTTAAGAAGCT	218						
Qy	270	CGATTTGCGCGCGTGAAGAAGATGCCACCGTTGATACCGCATAGCCAGAGTGGTG	329						
Db	219	AGATCTGGCAGCGTTAAAAAGGACCTGGCAGCCCTAATGACAGATTCACAGGACTGGTG	278						
Qy	330	GCCCGCTGACTGGGGGCACTACGGCGGTTTCATATCGTATGGCTGGCACTCCGCTGG	389						
Db	279	GCCAGCAGATTACGGTCATTATGGCCCTTCTTTATACGCATGGCTGGCACAGCGCCGG	338						
Qy	390	CACCTACCGTATTCGTGATGCCCTGGGGCGGTGGTACCGGAAGCCAGCGCTTTTGCACC	449						
Db	339	CACCTACCGTATTCGTGATGCCCTGGTGGCGGTGGCTCCGGCTCACAGCGCTTCGGCGCC	398						
Qy	450	GCTCAACTCTGGCGGCAACAGCTCAGCCTGGATTAAGCGCGCGCTCTGCTGTGGCCGAT	509						
Db	399	TCTCAATAGCTGGCCAGACAATGCCAATCTGGATTAAGACAGCTTGCCTCTTTGGCCCAT	458						
Qy	510	CAAGAAGAAGTACGGCAACAAAATCAGCTGGGCAGACCTGATGATCTTGGCTGGCACCCT	569						
Db	459	CAACAATAATACGGTCGAATAATCTCTGGCGGATCTAATGATCTACAGGNAACGT	518						
Qy	570	GGCTTATGAGTCCATGGGCTTACCTGCTTACGGCTTCTTTTCGGCCCGCTCGATATTG	629						
Db	519	AGCTCTGGAACATATGGCTTTAAACCTTTTGGTTTTGCAGGTGGCAGAGCAGATGTATG	578						
Qy	630	GGAAACCGAAAAGATATCTACTGGGTGAGCAAAAAGTGGCTGGCACCTTCTTGACGA	689						
Db	579	GGAGCCTGAAGAAGATGTATCTGGGAGCAGAAACCGAATGGGTGGGA-----GACAA	632						
Qy	690	ACGCTACGGCCAGCTGAACAGCCAGACCAATGGAAAACCGCTGGCGGCTGTGCCAAT	749						
Db	633	CGCTATGAAGGTACCCGAGAGC-----TCGAAATCCCTGGGAGCGGTACAAAT	683						
Qy	750	GGGTCTCATCTATGTGAACCGGAGGTGTTAAGCGGCCACCTGATCGCTGAGAACCGC	809						
Db	684	GGGACTCATCTATGTAAACCCGAGGACCCACAGCCAGCCAGCAGCCCTATCGCTGCTGC	743						
Qy	810	ACAGCAGGTACTTGAACCTTTCCGCCCTATGGCGATGAACGACGAAAACCGCAGCCCT	869						
Db	744	CGCTGATATCTGTCAGACTTTTGGCGAATGGCAATGAATCACGAAGAAACCGTGGCTCT	803						

Qy	1911	GACCGTCTCTGGCGGGTATGCGGTACTGGCGCACCACTATGTTGGCACCAACACGG	1970
Db	1884	GACTGCTTTGGTAGCGGGTATGCGGTACTGGCGCACCACTATGTTGGCACCACTG	1943
Qy	1971	CGTATTCACCGATTGTGAAGCGCCAGTTGACCAACGACTTTTGTGAACTGTACCGCATAT	2030
Db	1944	AGTGTTTACAAATAAGCGGGTCAGCTATCCAATGACTTCTTTGTAACCTGCTAGACCT	2003
Qy	2031	GGGACACGCTGG-----AAGCCGTAGTAGTACGACCGCTACGAATCCGGCACCGCAA	2084
Db	2004	CAACACTAAATGGCGAGCCAGCGATGAATCAGACAAAGTTTGTGAAGGCAGAGACTTCAA	2063
Qy	2085	GACCGGTGCGGTGAAGTGGACCCGCTCGCGGGTGGATCTGGTATTTGTGTTCCAACTCGCT	2144
Db	2064	AACGTGCGAAGTAAAGTGGAGTGGCACCCGGGTAGACTGATCTCGGATCCAAATCCGA	2123
Qy	2145	ACTTGGCGCTTTACGACAGAAGTGTACGCCAGGACGATAACGGCGGAGAAGTTCCTCAGAGA	2204
Db	2124	GCTAAGAGCCCTCGCAGAAGTGTACGGCTGTGCAGATTCCTGAAGAAAGTTTGTAAAGA	2183
Qy	2205	CTTCGTGGCGCCTGGACCAAGATGAACCGCCGCGCTTTCGA	2249
Db	2184	TTTTGTGAAGGCTCTGGCCCAAGTAAATGGACCTGGACCGGTTTGA	2228

RESULT 5
US-08-951-844-7

US 00 521 844 /
; Sequence 7, Application US/08951844
; Patent No. 6074860

APPLICANT: Robertson et al.

; TITLE OF INVENTION: Catalases
; NUMBER OF SEQUENCES: 8

NUMBER OF SEQUENCES: 0
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BA

ADDRESSEE: CAREER, DINNE, DE
ADDRESSEE: CECCHI, STEWART &
STREET: 6 BECKER FARM ROAD

STREET: O'LEARY HIGH ROAD
CITY: ROSELAND
STATE: NEW JERSEY

CITY: NEW JERSEY
COUNTRY: USA
ZIP: 07068

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETT

COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS

```

;
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/951
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/674,88
FILING DATE: July 3, 1996

ATTORNEY/AGENT INFORMATION:
NAME: Charles J. Herron

REGISTRATION NUMBER: 28,019
REFERENCE/DOCKET NUMBER: 3314

TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700

TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 7:

```

; INFORMATION FOR SEQ ID NO: 7:
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 2238 NUCLEOTIDES

```

LENGTH: 2238 NUCLEOTIDES
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE

STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: DNA

US-08-951-844-7

Query Match	29.7%	Score
Best [local] similarity	59.5%	Percentage

Best Local Similarity 39.38; File
Matches 1277; Conservative 0;

QY	150	CGGTACTTCCAAACAAGATTGGTGGCGGAAGGGTTGAACCTTGGATATTTTGCATCAGCA	209
DB	99	CGCGCAAAAACAGAGGATTGGTGGCCAACTGCTCAACCTCGGCTATTACGCCAACAA	158
QY	210	AGATCGCAATCAGACCCGATGATCCGGATTTCAACTACCGTGAAGAGTACGCAAGCT	269
DB	159	TTCATCGCTATCGACGCCAAACGACCGGATTTTGACTATGCGGAAGAGTTTAAGAACT	218
QY	270	CGATTTTCGACGCGCTGAAGAAAGATGTCCACGCGTTGATACCGCATAGCCAAAGATGGTG	329
DB	219	AGATCTGGCAGCGTTAAAAAGGACCTGGCGAGCGTAATCAGAGATTACAGAGCTGGTG	278
QY	330	GCCCGCTGACTGGGGGCACTACGGGGTTTGTATGATCCCGTATGCGTATGGCTATCCGCTGG	389
DB	279	GCCAGCAGATTACGGCTATTATGGCCCTTCTTTATACGCATGCGCTGGCACAGCGCG	338
QY	390	CACCTACCGTATTGCTGATCGCTGGGGCGGTGGTACCGGAAGCGACGCGCTTTCACC	449
DB	339	CACCTACCGTATCGGTGATGGCCGTGGTGGCGGTCCGGCTCAGCGCTTCGGCGCC	398
QY	450	GCTCAACTCTGGCCGGACAACTCAGCCTTGGATAAAGCGCGCGCTGCTGTGGCGCGAT	509
DB	399	TCTCAATAGCTGCCAGACAATGCCAATCTGATAAAGCAGCTTGTCTTTGGCCCAT	458
QY	510	CAAGAGAAGTACGGCAACAAAATCAGCTGGGAGAGCTGATGATCTGCGCTGGCACCGT	569
DB	459	CAAAACAAAATACGGTGGAAAATCTCCTGGCGGGATCTAATGATATCTACAGGAACGT	518
QY	570	GGCTTATGACTCATTGGGCTTACCTGCTTACGGCTTCTCTTTGGCGCGCTGCATATTG	629
DB	519	AGCTCTGGAAACTATGGGCTTTAAACCTTTTGGTTTTCAGAGTGGCAGACGATCTATG	578
QY	630	GGAAACCCGAAAAAGATATCTACTGGGTGACGAAAAAGAGTGGCTGGCACCTTCTGACGA	689
DB	579	GGAGCCTGAGAAGATCTATCTACTGGGAGCAGAAACCGAATGGCTGGGA-----GACAA	632
QY	690	ACGCTACGGCGAGTGAACAAGCCAGACAGACCATGGAAAAACCCGCTGGCGCTGTCCAAT	749
DB	633	CGCTATGAAGGTGACCCGAGAGC-----TCGAAAATCCCTGGGAGCCGCTACAAAT	683
QY	750	GGGTCTGATCTATGTGAACCCGGAAGGTGTTAACGGCACCCCTGATCCGCTGAGAACCCG	809
DB	684	GGGACTCATCTATGTAAACCCGAAGGACCCACGGCAAGCCAGACCTATPCGCTGCTGC	743
QY	810	ACAGCAGGTACTTTGAACCTTCGCCCCGATATGGGATGAAGCAGCAAAAAACCGCAGCCCT	869
DB	744	CGCTGATATCGTGAGACTTTTGGCCGAATGAAATGAATGACGAAGAAACCGTGGCTCT	803
QY	870	CACAGCTGGCGGCACACCGTGGTAAATGTACGGTAAATGGCAATGCCCTCTGCGT---T	926
DB	804	CATAGCGGGTGGACACACTTCGGAANAACCCATGGTGTCTGCCGATGCGGAGAAATATGT	863
QY	927	AGCCCTGACCCAAAAGCCCTGAGCTTGAACACCGAGGCTTAGGTTGGGCAACCCCAA	986
DB	864	GGGCCGAGACGCTGCCCGCGAGTATGAAGAAATGAGCCTGGGGTGGAAAAACACCTA	923
QY	987	CATGACGGGCAAGGCAACGCGCTGACCTCGGGTATCGAAGGTGCTTTGGACCCACAA	1046
DB	924	CGGCACCGACGCGTGGGATACCATCACCACTGGAGTGAAGAGCGGCTGGACCAAGAC	983
QY	1047	CCCCACGAAATTCGATATGGGCTATTTTCGACCTGCTGTTCGGCTACAAATTTGGGAAC	1106
DB	984	CCCTFACTCAATGGAGCAATAACTTTTGTGAAAACCTCTTTGGTTACAGTGGGAGCTTAC	1043
QY	1107	AAAGAGTCTTCGCGGTGCCCAACCATTTGGGAACCGATTGACATCAAAAAGGAAACAGAC	1166
DB	1044	CAAAAGTCCAGCTGGAGCTTATCAGTGGAAACCAAAAAGACGCGTCCCGGGGCTGCA	1103
QY	1167	GGTTGACGCCAGCACCCCTCTATTGCCACAACCCGATCATGACCGATGCGGATATGGC	1226
DB	1104	ACCGGATGCACATATCCAGCAGTGCAGCGCTCCATTTATGCTACTACGAGACCTGGC	1163
QY	1227	GATAAAGGTAAATCCGACCTATCCGGTATCTGCGAAAATTTATGCGCCGATCTCTGAGTA	1286

```

Db      1164  GTCGCCGATGGACCCCTGATTACGAAAAAATTTCTGACGGTACTATGAAACCCCTGATGA 1223
QY      1287  CTTTCAAGAAATTTTCGGAAGCGGTGTTCAAGCTGACGCCACCGTGAOCTGGGCCGCGAA 1346
Db      1224  GTTTCAGATGCTTTTCGCGAAGCATGGTACAACTGACACAGAGATATGGGACCAA 1283
QY      1347  ATCACTGTTACATCGGCCCGGAGTGCAGGAGAACCTGATTTGGCAAGACCCGATCC 1406
Db      1284  GGTGCGCTACCTGGGACCAAGAGTGCCTCAGGAAGACCTCATCTGGCAAGACCTATACC 1343
QY      1407  GCGAGTAAAC-----ACGACTACTCGGAGAGTGGTCAAGCAGAAATGTC 1454
Db      1344  AGATGTAGCCATCTCTTTGTAGACGAAACGATATTGAAGGCTTAAAGCCCAAAATCCT 1403
QY      1455  ACAAGTGGCCTGAGCATTAGTGAATGGTCTCCAGCGTTGGGACAGTGCCTGACTTAA 1514
Db      1404  GGATCGGAGTACCGGTAAAGGAGTGTAGACAGGATGGGCTTTCGATCTACTTT 1463
QY      1515  TCGCGGTTCCGATATCGCGGGGGTGTCTAACGGTGCCTGATTCGCTGGGCCCCACAGAA 1574
Db      1464  TAGAAATCTGCAAGCGCGGGTGCCTCAACGGTGCACGTATACGACTGGCCCAACAAA 1523
QY      1575  CGNCTGGCAGGCAAGCAGCGGAGCGCTGGGAAAGTGTGAG----- 1619
Db      1524  AGACTGGGAAGTAAACAAACCCCTCAGCAACTTCCAGGGTACTCAAAACACTAGAAGTAT 1583
QY      1620  -----CGTCTACGACGATCTCTGCCAGCACCGCGCTAGCATCGCGACGTGAT 1670
Db      1584  CCAGGAGGACTTTACCGAGCGCATCAGATACAAAGCAGTATCGTTGGCGGACCTGAT 1643
QY      1671  CGTCTGCGCGGTAGCGTATAGGATCGAGAAAGCGGAAAGCAGAGTGTACGATGGG 1730
Db      1644  TGTGCTGGCGCGGTGTGCGGGGTGTAGAAAAGCTGCAAAAGATGCTGGCCATGAGTGCA 1703
QY      1731  GFTTCCCTTCTGAAGGCGGTGGGATGCGGATGCGACCGCGAGATGACGACGACCTCTT 1790
Db      1704  GTGCGCTTTCAACCGGGAGGAGCGGTGCGCGCTGAGCAACCGATGTGGAGCTTT 1763
QY      1791  CGCACCGCTGGAGCCGCTGCGCGATGGCTTCGCCCACTGGCAGAGAAAGATGTGCT 1850
Db      1764  CGAAGCACTAGACCGAGCGCTGACGGCTTTAGAACTACATTAACCGGAGCATAACT 1823
QY      1851  GAAGCGGAGAGATGCTCTGATGCTGCGGAGTGTGGGCTTAACCGCGCGGAAAT 1910
Db      1824  ATCCGCTGAGGAAATGCTGTAAGCGGCGGAGCTTCTGCGCTTTCGGCACCAAGAAAT 1883
QY      1911  GACCGTGTGTCGGCGGTATGCGGTACTGGSCAACCACTATGGTGCCACCAACACGG 1970
Db      1884  GACTGTTTGTAGGCGGTATGCGGTACTGGCACCACTAGCAGGTTTCGCGATGCG 1943
QY      1971  CGTATTACCGATTTGTGAAGGCGATTTGACCAACGACTTTTGTGAACCTGACCGATAT 2030
Db      1944  AGTGTTTACAAATAAGCGGGTCACTATCCATGACTTCTTTGTAAACCTGCTAGACCT 2003
QY      2031  GGGGAACAGCTGG-----AAGCGGTAGGTAGCAAGCGCTACGAAATCCGCGACGCGAA 2084
Db      2004  CAACACTTAAATGGCGAGCGAGCGATGAATCAGACAAAGTTTGTGAAGGAGAGATTCAA 2063
QY      2085  GACCGTGTGCGGTGAGTGGACCGCTCGCGGTGATCTGTTATTTGGTTCCAACTCGCT 2144
Db      2064  AACTGGCGAAGTAAAGTGGAGTGGACCGCGGTAGACCTGATCTTCGGATCCAAATCCGA 2123
QY      2145  ACTGGCTCTTACCGAAGATGTAGCCCGAGGACGATACCGCGGAGAACTGCTGTCAGAGA 2204
Db      2124  GCTAAGAGCCCTCGCAGAGTGTACGGCTGTGCAGATTCTGAAGAAAGTTTGTAAAGA 2183
QY      2205  CTTGCTGCGCGCTGGACAAAGTATGAACCGCGACCGTTTCGA 2249
Db      2184  TTTTGTGAAGGCGTGGGCGAAAGTAAATGACCTGGACCGGTTTGA 2228

```

```

US-09-412-347-7
; Sequence 7, Application US/09412347
; Patent No. 6410290
; GENERAL INFORMATION:
; APPLICANT: Robertson, Dan E.
; APPLICANT: Sanyal, Indrajit
; APPLICANT: Adhikari, Robert S.
; TITLE OF INVENTION: CATALASES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/412,347
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/674,887
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Halle, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09015/002001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2238 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...2235
US-09-412-347-7

Query Match      29.7%; Score 672.2; DB 4; Length 2238;
Best Local Similarity 59.5%; Pred. No. 3.7e-179;
Matches 1277; Conservative 0; Mismatches 808; Indels 60; Gaps 6;

QY      150  CGGTACTTCCAAACAAGATTGGTGGCCGAAGGGTTGAACCTGGATATTTTGCATCAGCA 209
Db      99  CGGCACCAAAACAGGATTGGTGCCCAACATGCTCAACCTCGCATCTTACGCCAACA 158
QY      210  AGATCGCAANTCAGACCCCGATGATCGGATTTCAACTACCTGCTGAGAGAAGTACGCAAGT 269
Db      159  TTTATGCTGCTATCGGACCCCAAGACCGGATTTTGAATATGCGAAGAGCTTTAAGAAGCT 218
QY      270  CGATTTTCAGCGCTGAAGAAAGATGTCACACCGTTTGATGACCGATACCGCAAGTGGTG 329
Db      219  AGATCTGGCAGCGGTTTAAAGAGGACCTGGCAGCGCTAATGACAGATTCACAGGACTGGTG 278
QY      330  GCCCGCTGACTGGGGGCACCTAGCGGGGTTTGTATGATCCGTATGGCTTTGGCACTCCGCTGG 389
Db      279  GCCAGCAGATTACGGTCAATTATGGCCCTTCTTTATACGATGGCGTGGCAGCGCCGG 338
QY      390  CACCTACCGTATGCTGATGGCGGCGGCGGTTGATCCGGAAGCAGCGCTTTTGCAC 449
Db      339  CACCTACCGTATGCTGATGGCGGCGGCGGTTGATCCGGAAGCAGCGCTTTTGCAC 398
QY      450  GCTCAACTCTCGCGGCAACAGTCAAGCTGATGAAGCGCGCTCTCTGCTGTCGCGCAT 509
Db      399  TCTCAATAGCTGGCCAGACATGCCAANTCTGGATAAGCAGCTTGTCTTTTGGCCCAT 458

```

Qy	510	CAAGAAGAGTACGGCAACAAAAATCAGCTGGGCGAGACCTGATGATTCTGGCTGGCACCGT	569
Db	459	CAAAACAAAAATACGGTGGAAAAATCTCTGGGCGGATCTAAATGATACTACAGGAACGT	518
Qy	570	GGCTTATGAGTCATCGGCTTACCTGCTTACGGCTTCTCTTTTCGGCGCGGTGATATTG	629
Db	519	AGCTCTGGAACATATGGCTTTTAAACTTTTGGTTTGCAGGTGGCAGAGCATGTATG	578
Qy	630	GGAACCCGAAAAAGATATCTACTTGGGTGACGAAAAAGAGTGGCTGGCACCTTCTTGACGA	689
Db	579	GGACCTGGAAGAAGATGATATCTACTTGGGAGCAGAAACCGAATGCTGGGA-----GACAA	632
Qy	690	ACGTCACGGCGAGTGAAACAAGCCAGAGACCATTGGAAAAACCCGCTGGCGGCTCTCAAAAT	749
Db	633	CGCTATGAAGTGCACGAGAGC-----TCGAAAAATCCCTGGGAGCCCTGACAAAT	683
Qy	750	GGGTCTGATCTATGTGAACCCGGAAGGTGTTACGGCCACCCCTGATCCGCTGAGAACCCG	809
Db	684	GGGACTCATCTATGTAAACCCGGAAGGACCAACGGCAAGCCAGACCCCTATCGCTGCTGC	743
Qy	810	ACAGCAGGTACTTTGAAACCTTCGCCGATGCGGATGAACGACGAAAAAACCCGACCCCT	869
Db	744	CGGTGATATTCTGTGAGACTTTGGCCGAATGGCAATGAATGACGAGAAACCGTGGCTCT	803
Qy	870	CACAGTGGCGGCCACACCGCTCGGTAAATGTACGGTAATTGGCAATGCCCTCTCGCT---	926
Db	804	CATAGCGGTGGACACACCTTCGGAAAAACCCATGGTGTGCCGATCGGAGAAATATGT	863
Qy	927	AGCCCTTGACCCAAAAGCCTCTCAGCTTGAACACGAGGCTTAGTTGGGGCAACCCCAA	986
Db	864	GGGCCGAGAGCCCTGCCGCCGAGGTATGAAGAAATGAGCCTGGGTTGGAAAAACACCTA	923
Qy	987	CATGCAAGGCAAGCAACGCGGTGACTCGGTTATCGAAGTGTCTGGGACCAACCA	1046
Db	924	CGGCACCGACGGTCCGATACCATCACCATGAGCTAGAGGCGCTGGACCAAGAC	983
Qy	1047	CCCCACGAATTCGATATGGCTATTTTCGACCTGCTGTTTCGGGTACAAATTTGGAACTGA	1106
Db	984	CCCTACTCAATGGAGCAATAACTTTTGTGAAAAACCTCTTTGTTAGAGTGGGAGCTTAC	1043
Qy	1107	AAGAGTCTTCGGGTGCCACCATTTGGGAACGATTGACATCAAAAGGAAACCAAGCC	1166
Db	1044	CAAAAGTTCAGCTGGAGCTTATCAGTGGAAACCAAAAGAGCGGTGCCGGGCTGGCACCAT	1103
Qy	1167	GGTTGACGCGAGGACCCCTCTATTGCCACAACCCGATCATCACCGATCGGATATGGC	1226
Db	1104	ACCGATGCATGATCCAGCAAGTCGCACGCTCCATTTATGCTCACTACGACCTGGC	1163
Qy	1227	GATNAGGTAAATCCGACCTATCGCGCTATCTCGCAAAATTCATGCCGATCCTGAGTA	1286
Db	1164	GCTCGCATGGACCTGATTACGAAAAAATTTCTCGACGGTACTATGAAAAACCTGATGA	1223
Qy	1287	CTTCAAGAAACCTTCGGAAGCGGTGGTTCAAGCTGACGACCGTACCTGGGCGCGAA	1346
Db	1224	GTTTTCAGATGCTTTTCGGAAGCATGGTACAACTGACACAGAGATATGGGACCAA	1283
Qy	1347	ATCACGTTTACATCGCCCGGAAGTGGCGGAGAGACCTCATTTTGGCAAGCCGATTC	1406
Db	1284	GGTCGGTACTGGGACCAAGTGCCTCAGGAAGACCTCATCTGGCAAGCCCTATACC	1343
Qy	1407	GGCAGGTAAC-----ACCGACTACTCGAGAAAGTGGTCAAGCAGAAAAATTCG	1454
Db	1344	AGATGTAAGCATCTCTTGTAGACGAAAAACGATATTGAAGGCCTAAACCCCAAAATCCT	1403
Qy	1455	ACAAAGTGGCTGAGCAATTAGTGAGATGGTCTCCACGGTTGGGACAGTGGCCCGTACTTAA	1514
Db	1404	GGATCGGAGTCAACGCTAAGACGAGCTGCTAAGCACGGCATGGGCTCTGCATCTCTTT	1463
Qy	1515	TCGGGTTCCGATATGCGCGCGGTGCTAAACGTTGCCCGATTCGCTTGGCCCCACAGAA	1574
Db	1464	TAGAACTCTGACAAGCGCGGTGCCAACGGTGCACGTATACGACTGGCCCCACAAA	1523

Qy	1575	CGAGTGGCAGGGCAACGAGCGCGAGGCGCTGGCGAAAGTGTCTGAG-----	1619
Db	1524	AGACTGGGAAGTAAACAACACCTCAGCAACTTGCAGGGTACTCAAACACTAGAAGTAT	1583
Qy	1620	-----CGTCTACGAGCAGATCTCTGCCGACACCGCGCTAGCATCGCGACCTGAT	1670
Db	1584	CCAGGAGGACTTTTAAACCAGCGCAATCAGATACAAGCAGTATCTTTGGCCGACCTGAT	1643
Qy	1671	CGTTCTGGCCGGTAGCGTAGGCATTCGAGAAAGCCGCGAAAGCAGCAGGTATACGATGGCG	1730
Db	1644	TGTGCTGGCCGGCTGTGCGGGTGTAGAAAAGCTGCAAAAGATGCTGGGCATGAGGTGCA	1703
Qy	1731	CGTTCCCTTCTTGAAGAGCCGTGCGCATCGACCGCGGAGATCAGCCAGCAGCAGACTCCT	1790
Db	1704	GTTGCCCTTTCAACCCGGGACACCGGATGCCACCGCTGAGCAAAACGATGTGGAACTTT	1763
Qy	1791	CGCACCGCTGGAGCCGCTGCCGATGGCTTCCGCAACTGSCAGAGAAGAGTATGTGT	1850
Db	1764	CGAAGCACTAGCAGCGGCTGACGGCTTTAGAACTACTATTTAAACCGGAGCATAAAGT	1823
Qy	1851	GAAGCCGGAAGAGATGCTGCTGGATCGTGGCAGCTGATGGGCTTAAACCGGCCGAAAT	1910
Db	1824	ATCCGCTGAGGAATATGCTGATACCGGGCGCAGCTTCTGTCTGCTTTTCGGCACAGAAAT	1883
Qy	1911	GACCGTGCTCTGGCGGGTATGGCGGTACTGGGCACCAACTATGTGGCACCAAAACACGG	1970
Db	1884	GACTGCTTTGTTAGGCGGTATGCTGTACTGGGCACCACTACGACGGTTCGCAGATGG	1943
Qy	1971	CGTATTCACCGATTGTGAAGCCAGTTTGACCAACGACTTTTGTGAACCTGACCGATAT	2030
Db	1944	AGTGTTTACAANTAAAGCCGGTGCAGCTATCCATGACTTCTTTGTAACCTGTGACCT	2003
Qy	2031	GGGGAACAGCTGG-----AAGCCGTTAGTACCAACGCTACGAAATCGCGACCGCAA	2084
Db	2004	CAACACTAATGGCAGCCAGCGATGAATCAGACAAGATTTTGAAGGCAGAGACTTCAA	2063
Qy	2085	GACCGTGCCTGAAAGTGAGCCGCTCGCGGGTGGATCTGGAATTTGTGTTCCAACTCGCT	2144
Db	2064	AATCGGCGAAGTAAAGTGGAGTGGCACCAGCCGGGTAGACCTGATCTCGGATCCCAATCCGA	2123
Qy	2145	ACTGCGCTCTTAGCCAGAAGTGTACGCCAGGACGATAACGGCGAGAAGTTCCCTCAGAGA	2204
Db	2124	GCTAAGAGCCCTCGCAGAAGTGTACGGCTGTGCAGATCTGAAGAAAGATTTGTTAAGA	2183
Qy	2205	CTTTCGTCCGCGCTGGACCAAGATGATGAACCGCGACCGTTTCGA	2249
Db	2184	TTTTGTGAAGCCTTGGCCCAAGTAAATGGACCTTGGACCGGTTTGA	2228

```

RESULT 7
US-08-418-782-20
: Sequence 20, Application US/08418782
: Patent No. 5658733
: GENERAL INFORMATION:
: APPLICANT: Cockerill, Franklin R.
: APPLICANT: Kline, Bruce C.
: APPLICANT: Uhl, James R.
: TITLE OF INVENTION: Detection of Isoniazid Resistant Strains
: TITLE OF INVENTION: of M. Tuberculosis
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Schwegman, Lundberg & Woessner
: STREET: 3500 IDS Center
: CITY: Minneapolis
: STATE: MN
: COUNTRY: USA
: ZIP: 55402
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/418,782

FILING DATE: 435
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION: Woessner, Warren D.
REGISTRATION NUMBER: 30,440
REFERENCE/DOCKET NUMBER: 150.141US1
TELECOMMUNICATION INFORMATION: 612-339-0331
TELEPHONE: 612-339-3061
TELEFAX: 612-339-3061
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 2331 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 70...2289
US-08-418-782-20

Query Match 27.3%; Score 618.6; DB 1; Length 2331;
Best Local Similarity 58.7%; Pred. No. 4.6e-164;
Matches 1253; Conservative 0; Mismatches 819; Indels 63; Gaps 8;

160 AACAAAGATTGGTGGCGGAGGTTGAACCTGGATATTTTGCATGACGACGATCGCAAA 219
172 AACAGGACTGGTGGCGCAACCGGCTCAATCTGAAGGTACTGACCAAAACCGCGCGTC 231
220 TCAGACCGGATGGATCGGATTTCAACTACCGTGAAGAAGTACCAAGCTCGATTTCGAC 279
232 GCTGACCGGATGGTGGCGGCTTCGACTATGCCCGGAGGTGCGGACCATCGAGCTTGAC 291
280 GCGCTGAAGAAAGATGTCACGCGCTTGATGACCGATAGCCAAAGAGTGTGGCGCGTGAC 339
292 GCCCTGACGGGACATCGAGGAAGTATGACCACTCGACGCGGTGGCGCGCGCGAC 351
340 TGGGGGACATACGCGGTTTGTATGATCCGTTATGCTGGTGGTGGTGGTGGTGGTGGTGG 399
352 TACGGCCACTACGGCGCGCTGTTTATCCGGATGGCGTGGCAGCGTGGCGCGCTTACGAC 411
400 ATTCTGATGGCGTGGCGGCGTGGTACCGGAGCCAGCGCTTGGCACCGCTCACTCC 459
412 ATCCACAGCG 471
460 TGGCGGACAACTGACGCTGGATAAAGCGCGCGCTGCTGTGGCGGATCAAGAAGAG 519
472 TGGCGGACAACTGACGCTGGATAAAGCGCGCGCTGCTGTGGCGGATCAAGAAGAG 531
520 TAGGGCAAAATACGCTGGGAGACCTGATGATTCGCTGGCAGCGCTGGCAGCGCTTATGAG 579
532 TAGGGCAAAAGCTCTCGGGGAGACCTGATTTTTCGCGGCAACTGCGCGCTGGA 591
580 TCCATGGGCTTACCTGCTTACGCTTCTCTTTCGCGCGCTGCTGATTTTGGGAACCGGAA 639
592 TCGATGGGCTTCAAGAGCTTCGGGTTCGGGTTCGGCGGCTGACCACTGAGTGGGAGCGCG 649
640 AAGAGATATCTACTGGGTGACGAAAGAGTGGCTGGCAGCTTCTGACGAAACGCTACGGC 699
650 -ATGAGGTCTATTGGGGCAAGGAGCCACCTGGCTG- - - - -CGGATGAGCGTTACAGC 702
700 GAGGTGAACAGCCAGACCATGGAAACCGCTGGCGGCTGTGCAAAATGGGTCTGATC 759
703 GGTAAGGGGATC- - - - -TGGAGAACCCGCTGGCGGCTGACAGATGGGCTGATC 753
760 TATGTGAACCCGGAAGTGTTAAGGGCCACCTGATGCGGTGAGAACCCGACAGCAGGTA 819
754 TACGTGAACCCGAGGGGGGCAACCGCAACCCGAGCCCATGCGCGCGGCGGCTGACATT 813
820 CTTGAAACCTTCGCGCGCTATGGGATGAACGAGCAAAACCCGACCGCTTCACAGCTGGC 879
814 CGCGAGACGTTTCGGCGCATGGCCATGAACGAGCTGCAAAACAGCGGCGCTGATCGTGGC 873

880 GGCACACCGCTCGGTAAATTTGACGGTAAATGCAATGCTCTGCTGCTAGCCCTGACCCA 939
874 GGTACACTTTCGCTAGAGCCATGGCGCGCGCGCGCGCTGCTGCTGCGCGCGGACCC 933
940 AAGCCCTCTGACGTTGAAACACAGGCTTAGTTGGGGCAACCCCAACATGACAGGCAAG 999
934 GAGGTGCTCGCTGGAGCAGATGGGCTTGGGCTTGAAGAGCTGCTGATGGCACCGGAACC 993
1000 GCAAGCAACCGCTGACCTCGGGTATCGAAGGTGCTTGGACCACCAACCCACCAAAATTC 1059
994 GGTAAAGACCGGATCACACCGCATCGAGTCTGATGGAGCAACACCCGACGAAATGG 1053
1060 GATATGGCTATTTTCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1119
1054 GACAAAGTTTCTCGAGATCTGACGGCTACAGTGGAGTGGAGTGGAGAGGCGCTGCT 1113
1120 GGTCCCAACCATGCGGACCGATTTGACATCAAAAGAAAGAAACAGCGGTTGACCGCAGC 1179
1114 GGCCTTTGGCAATACACCGCAAGGAGGCGCGCTGCGCGCACCATCCCGGACCGCTTC 1173
1180 GACCCCTCTATTTCGCGCAACCCGATGACCGATGCGGATATGCGGATATAAGGTAAAT 1239
1174 GCGCGGCT- - -AGGCGCTCCCGACGATGCTGCGCACTGACCTCTCGCTGCGGTGGAT 1230
1240 CCGACCTATCGCGCTATCTGCGAAAAATTTATGCGCGCTGCTGAGTACTTTCAAGAAAACT 1299
1231 CCGATCTATGAGCGGATCACGCGTCTGCTGCTGCGCAACCCCGAGGATTTGGCGGACGAG 1290
1300 TTCGGAAGGGTGGTTCAGAGCTACGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1359
1291 TTCGCAAGGCTGGTACAAAGTATCCACGAGACATGGGTCTGCTGCTGCTGCTGCTGCTGCT 1350
1360 GCGCGGAGTGGCGGCGAGAGACCTGATTTGGCAAGACCGGATTTCCGCGAGGTAACACC 1419
1351 GGGCGGTGCTCCCAAGCAGACCTGCTGTGGCAGGATCCGCTCCCTGCGGTGAGCGAC 1410
1420 GACTACTGCGAAGAGTGG- - - - -TCAAGCAGAAAAATTTGACAAAAAGTGGCGCTG 1467
1411 GACCTGCTGCGAAGCGGAGATTGCGCAGCTTAAAGAGCCAGATCCCGGATCGGATG 1470
1468 AGCATTAAGTATGATGCTCCACCGCTTGGACAGTCCCGTACTTATCCGCGGTTCGAT 1527
1471 ACTGCTCAGAGCTAGTTTCGACCGATGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1530
1528 ATGCGCGCGCTGCTAACGCTGCGCGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1587
1531 AAGCG 1590
1588 AAGAGCGCGA- - -GCGCTTGGCGAAGTCTGAGCGCTTACGAGCAGATCT- - - - - 1636
1591 AAGACCG 1650
1637 - - - - -CTGCGCGACCGCGCGCTAGCATCCGCGAGTATGCTGCTGCTGCTGCTGCTGCTGCT 1680
1651 TTCAACTCCG 1710
1681 GGTAGCTAGGCTACGAGAAAGCGCGGAAAGCAGCAGCTTACGATGCTGCGGTTCCTCCTC 1740
1711 GGCTGCTGCG 1770
1741 CTGAAGCGCGCTGGCGATGCG 1800
1771 ACCCGGCG 1830
1801 GAGCGGTGCG 1860
1831 GAGCGCAAGCAGATGCTTCCGAAATACCTCGGAAAGGCGCAACCGCTTCCGCGCGCGAG 1890
1861 GAGATGCTGCTGATGCTGCGAGCTGATGGCTTAAACCGCGCGCGCGCGCGCGCGCGCGCG 1920
1891 TACATGCTGCTGCAAGGCGAACCCTGCTTACGCTACGTCCCTCGATGATGAGTGGGTGCTG 1950

QY 1921 CTGGCGGTATGCGGTACTGGGACCAACTATGTTGGCACCACCAACACAGCGGTATTTCACC 1980
Db 1951 GTAGTGGCCCTCGCGGTCTCGCGCAAACTACAAAGCGCTTACCCTCGCGGTGTTCACC 2010
QY 1981 GATTGTGAAGCCAGCTTGACCAACACACTTTTGTGTAACCTGACCGGATATGGGAAACAGC 2040
Db 2011 GAGGCTCGGAGTCACTGACCAACGACTCTCTGTGAACCTGCTGACATGGGTATACC 2070
QY 2041 TGGAAAGC---CGGTAGGTAGCAACGCCCTACGAAATCCCGACCGCAAGACCGGTCCCGTG 2097
Db 2071 TGGGAGCCCTCGCCAGCAGATGACCGGACCTACCAAGGCAAGGATGGCAGTGGCAAGGTG 2130
QY 2098 AGTGGACCGCTCGCGGTGGATCTGTTATTTGTTTCCAACTCCTACTCGCTTTCAC 2157
Db 2131 AAGTGGACCGCGAGCGGTGGACCTGTCTTCGGGTCCAACTCGCGAGTTGCGGCGCTT 2190
QY 2158 GCAGAAGTGTACGCCCGCAGGACGATAAACGGCGGAGAAGTTTCGTGAGAGTTTCGTGCGCGCC 2217
Db 2191 GTCGAGGTCATGCGCGCGGATGACGCGGACGCGAAGTTTCGTGCGAGGACTTTCGTGCGCGCC 2250
QY 2218 TGGACCAAGTGAAGCAACCGCGACCGTTTTCGACGT 2252
Db 2251 TGGGCAAGGTGATGAACCTCGACAGGTTTCGACGT 2285

RESULT 8

US-08-852-219-20
; Sequence 20, Application US/08852219
; Patent No. 5922575
; GENERAL INFORMATION:
; APPLICANT: Cockerill, Franklin R.
; APPLICANT: Kline, Bruce C.
; TITLE OF INVENTION: Detection of Isoniazid Resistant Strains
; TITLE OF INVENTION: of M. Tuberculosis
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Muecking, Raasch & Gebhardt, P.A.
; STREET: 119 No. 5922575th Fourth Street, Ste. 203
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,219
; FILING DATE: 07-May-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sandberg, Victoria A.
; REGISTRATION NUMBER: 41,287
; REFERENCE/DOCKET NUMBER: 230.00010130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1226
; TELEFAX: 612-305-1228
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2331 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 70..2289
US-08-852-219-20

Query Match 27.3%; Score 618.6; DB 2; Length 2331;
Best Local Similarity 58.7%; Pred. No. 4.6e-164;

Matches 1253; Conservative 0; Mismatches 819; Indels 63; Gaps 8;
QY 160 AACAAAGATTGGTCCCGAAGGGTTGAACCTTGGATATTTTTCATCAGCAAGATCGCAAA 219
Db 172 AACAGGACTGGTGGCCCAACCGCTCAATCTGAAGGTACTGCACCAAAACCGGCCGTC 231
QY 220 TCAGACCCGATGGATCCGGATTTCACTACCGTGAAGAGTACGGAAGTCTGATTTTCAG 279
Db 232 GCTGACCCGATGGTTCGCGCTTCGACTATGCGCGGAGGTGCGACCATCAGCGTTGAC 291
QY 280 GGCCTGAAGAAAGATGTCACCGCTTGATGACCGGATAGCAAGAGTGGTGGCCCGCTGAC 339
Db 292 GCCCTGACCGGGACATCGAGGAAGTATGACCACTCGCAGCGGTGGTGGCCCGCGAC 351
QY 340 TGGGGCACTACGGCGGTTTGTATGATCCGTATGGCTTGGCACTCGCGTGGCACTACCGT 399
Db 352 TACGCCCACTACGGCGCGCTGTTATCCGGATGGCTGGCAGCTGCGCGCACTACCGC 411
QY 400 ATTGCTGATGGCCGTGGGGCGGTGATACCGGAAGCAGCGCTTTTGACACCGTCAACTCC 459
Db 412 ATCCACGACGGCG 471
QY 460 TGGCGGCAACAGCTCAGCTGATAAAGCGCGCGCTCTCTCTGTCGCCGATCAAGAGAAG 519
Db 472 TGGCGGCAACAGCTGATGACAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 531
QY 520 TACGGCAACAAATCAGCTGGGCGACAGCTGATGATTTCTGGCTGGCAGCGCTGCTATGAG 579
Db 532 TACGGCAAGAGCTCTCATGGCGGCACTGATTTTTCGCCGCACTGCGCGCTGGAA 591
QY 580 TCCATGGGCTTACCTGCTTACGGCTTCTCTTTTCGCCGCGCTGATTTTGGAAACCGGAA 639
Db 592 TCGATGGGCTTCAAGACGTTTCGGTTTCGGCTTCGGCGCGGTGCGACAGTGGAGCGCG 649
QY 640 AAGATATCTACTGGGTGACCAAAAGAGTGGCTGGCAGCTCTTGACAAAGCTACCGC 699
Db 650 -ATGAGGCTATTGGGGCAAGGAGGACCTGCGT-----GGATGAGGCTTACAGC 702
QY 700 GACGTGAACAGCCAGAGACCACTGGAACACCGCGCTGGCGCTGTCCAAATGGGTCTGATC 759
Db 703 GGTAAAGCGGATC-----TGGAGAACCGCTGGCGCGGTGCAGATGGGCTGATC 753
QY 760 TATGTGAACCCGGAAGGTGTTAAGCGCCACCTGATCCCTGAGAACCCGACAGAGGTA 819
Db 754 TACGTCAACCCGAGGGCGCGCAACCGCGACCCCATGGCGCGCGGTGCGACAT 813
QY 820 CTTGAACCTTCGCCGTATGGGATGAACGACGAAAAACCGCAACCGCTCACAGTGGC 879
Db 814 CCGGAGCGTTTCGGCGCATGGCCATGAACAGCTGGAACACGCGCGCTGATCTCGCGC 873
QY 880 GGCCACACCGTCGGTAAATGTCACGTAATGGCAATGCGCTCTGCGTTAGCCCTGACCCA 939
Db 874 GGTACACTTTCGTTAAGACCCATGGCGCGCGCGCGCGCGATCTGTCGCGCGCGAACCC 933
QY 940 AAGGCTCTGACGTTGAAACCAAGGCTTAGGTTGGGCAACCCCAACATGACGAGGCAAG 999
Db 934 GAGGCTCTCGCTGGAGCAGATGGCTTGGCTTGGAGAGCTCGTATGGCACCGGAACC 993
QY 1000 GCAAGCAACCGCTGACCTCGGATATCGAAGGTGCTTGGACCAACCAACCGCGCAAAATTC 1059
Db 994 GGTAAAGACGCGATCACCAGCGCATCGAGGTCTGATGGACGAACACCGCGCAAAATGG 1053
QY 1060 GATATGGGCTATTTGACCTGCTGTTTCGGCTACAATTTGGGAAGTGAAGAGTCTCTGCC 1119
Db 1054 GACAACAGTTTCTCGAGATCTCTGACGCTACGAGTGGAGGTGACGAGGAGCGCTTCT 1113
QY 1120 GGTGCCCACTATGGGAACCGATTCGATCAAAAAGGAAACCAAGCGGTTGACGCGCAG 1179
Db 1114 GCGCTTGGCAATACACCCCAAGGACGCGCGGTGCGCGCACTCCGCGACCGGCTTC 1173
QY 1180 GACCCCTCTATTCGGCAACACCGGATGACCGGATGCGGATGCGGATATGAAGGTAAAT 1239
Db 1174 GCGGGGCC---AGGGCGCTCCCGCGAGATGCTGGCCACTGACCTCTCGCTGCGGCGTGGAT 1230

Qy	1240	CCGACCTTATCGCGTATCTCGGAAAATTTATGATGGCCGATCCTGAGTACTTTCAAGAAACT	1299
Db	1231	CGGATTTATGAGCGGATTCACGCGTCTGCTTGGAAACACCCGAGGAATTTGGCGACGAG	1290
Qy	1300	TTTCGGAAGCGTGGTTTCAAGCTTGACGACACCGTGACCTGGGCGCGAAATCACGTTTACATC	1359
Db	1291	TTTCGCAAGCGTGGTTTCAAGCTTGATCCACCGAGACATGGGTCCGTTGGCAGATACCTT	1350
Qy	1360	GGCCCGGAATGGCCGCGAAGAGCTGATTTTGGCAAGACCCGATTCGGCAGGTAAACACC	1419
Db	1351	GGCGCGTGGTCCCAAAGCAGACCCCTGCTGTGGCAGGATCCGCTCCCTCGGTCAGCCAC	1410
Qy	1420	GACTACTGCGAAGAAGTGG-----TCAAGCAGAAAAATTGCACAAAAGTGGCGCTG	1467
Db	1411	GACTCTCGCGGAAGCCGAGATGTCGACGCTTTAAGAGCCAGATCGGGGCATCGGGATG	1470
Qy	1468	AGCATTTAGTGAATGGTCTCCACCGCTTGGGACAGATGCCCGTACTTTATCCGGTTCGCGAT	1527
Db	1471	ACTGTCTCACAGCTAGTTTCGACCGCATGGCGGCGGCGTCTGTTCCGTGGTAGCGAC	1530
Qy	1528	ATGGCGGGCGGTGCTTAACGTTGCCCGCATTCGGTTTGGCCCCCACAAGACGAGTGGCAGGGC	1587
Db	1531	AAGCGCGGCGGCCAACAGGTGTGCGATCCGCTTCGAGTCACAAGTTCGGGTGGGAGTTC	1590
Qy	1588	AAGCAGGCCGA--GCSCCTTGGCGAAAGTCTGAGGCTTACGAGCAGTCT-----	1636
Db	1591	AAGCAGCCCGACGGGATCTTCCGAGGTCATTTTCGACCCCTGGAGAGATCCAGGAGTCA	1650
Qy	1637	-----CTGCCGACACGGCGCTAGCATCCGCGACGCTGATCGTTCTTGGCC	1680
Db	1651	TTCAACTCCCGCGCGCGGGAACATCAAAAGTGTCTTTCGCGACCTCGTCTGCTCGGT	1710
Qy	1681	GSTAGCCTAGGCATCGAAGAAGCCGCAAGACGACAGGTTTACGATGTGCGCTTCCCTTC	1740
Db	1711	GGCTGTGCCGCCATAGAGAAGCAGCAAGCGCGCTTGGCCACAACATCAACGTTGCCCTTC	1770
Qy	1741	CTAAAGGCCGTGGCATGCGACCGCGGAGATGACCGACGACGACATCTTTCGCGACCGCTG	1800
Db	1771	ACCCGGCGCGCACGGATGCTCGCAGGAACAACCGACGTGGAACTCTTTCGCTGCTG	1830
Qy	1801	GAGCGCTGGCCATGGCTTCCGCAACTGGCACAAGAAAGATGTGTGTGAAGCCGGAA	1860
Db	1831	GAGCCCAAGGCATGGCTTCCGAAACTACCTCGGAAGGGCAACCCGTTGCCGCGCGAG	1890
Qy	1861	GAGATGCTGTGGATGTGCGCAGCTGATGGGCTTAAACGGCCCGGAATGACCGTGTCTG	1920
Db	1891	TACATGCTGTGCACAAGGCGAACCTTCTTACGCTCAGTGCCTCTGATGACGGTGTCTG	1950
Qy	1921	CTGGCGGGTATGGCGTACTTGGGCACCAACTGTGTGGCAACCAACGCGGTATTTCACC	1980
Db	1951	GTAGTGGCGTGGCGCTCTCGCGCAAACTACAAGCGCTTACCCGTGGGCGCTGTTCACC	2010
Qy	1981	GATTTGTGAAGGCCAGTTGACCAACGACTTTTGTGAACCTGACCCATATGGGGAACAGC	2040
Db	2011	GAGCCTCCGAGTCACTGACCAACGACTTCTCTGTAACCTGCTCCACATGGGTATCAC	2070
Qy	2041	TGGAAG----CGTAGGTAGCAACGCCTACGAAATCCGCGACCGCAAGACCGGTGCGGTG	2097
Db	2071	TGGGAGCCCTTCGCCACGACATGACGGGACCTTACCAGGGCAAGGATGGCAGTGGCAAGGTG	2130
Qy	2098	AAGTGCACCCCTCGGGGTGGATCTTGTTATTTGGTTCCAACTCGCTACTTGGCTCTTAC	2157
Db	2131	AAGTGCACCGGCGAGCGGTGGACCTTCTCGGGTCCAACCTCGGAGTTGGGGCGCTT	2190
Qy	2158	GCAGAAGTGTACGCCCAGGACGATAACGGCGAGAAGTTTCGTTCAGAGACTTTCGTCCGCCGC	2217
Db	2191	GTCCAGGTCTATGGCGCCGATGACGCGCAGCCGAACTTCGTGCAGGACTTCGTCTCGTGC	2250
Qy	2218	TGCACCAAGTGTATGAACGCCCGACCCGTTTCGAGCT	2252
Db	2251	TGGGACAAGGTGATGAACCTTCGACAGTTTCGAGCT	2285

```

RESULT 9
US-09-103-840A-2/c
: Sequence 2, Application US/09103840A
: Patent No. 6294328
: GENERAL INFORMATION:
: APPLICANT: FLEISCHMAN, Robert D.
: APPLICANT: WHITE, Owen R.
: APPLICANT: FRASER, Claire M.
: APPLICANT: VENTER, John C.
: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
: TITLE OF INVENTION: TUBERCULOSIS
: FILE REFERENCE: 24366-20007.00
: CURRENT APPLICATION NUMBER: US/09/103.840A
: CURRENT FILING DATE: 1998-06-24
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 2
: LENGTH: 4403765
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis
: FEATURE:
: OTHER INFORMATION: CDC 1551
: OTHER INFORMATION: "n" bases at various positions throughout the sequence
: OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

Query Match	27.3%	Score 618.6	DB 3	Length 4403765
Best Local Similarity	58.7%	Pred. No. 1.1e-162		
Matches 1253	Conservative	0	Mismatches 819	Indels 63
Gaps	8			
QY	160	AACAAGATGGTGGCCGAAGGTTCAACCTGGATATTTTGCATCAGCAAGATCGCAAA	219	
Db	2153267	AACAGGACATGGTGGCCCAACCGGCTCAATCTGAAGGTACTGCACCAAAACCCGGCGTC	2153208	
QY	220	TCAGCCCGATGGATCCGGATTTCAACTACCGTGAAGAAGTACGCAAGCTCGATTTTCGAC	279	
Db	2153207	GCTGACCCGATGGTGGCGGCTTCGACTATGCCGGAGGTCGGACCATCGAGGTGAC	2153148	
QY	280	GCCTGAAGAAAGATGTCCACGGGTTGATGACCGATAGCCAAAGTGGTGGCCCGGTGAC	339	
Db	2153147	GCCTGACGGCGGACATCGAGGAAGTATGACCACCTCGCAGCGCTGTTGGCCGCGGAC	2153088	
QY	340	TGGGGGCATACGGCGGTTTATGATCCGTATGGCTTGGCACTCGCGTGGCACTACCGT	399	
Db	2153087	TAGGCCACCTACGGCGGCTGTTTATCCGATGGGTGGGACGCTGCCGACCTACCGC	2153028	
QY	400	ATTGCTGATGGCGTGGGGCGGTGGTACCAGGAGCCAGCGCTTGCACCGCTCAACTCC	459	
Db	2153027	ATCCAGCAGCGCGCGCGCGCGGGGGGCGATGCAGGGTTGCGCGCGCTTAACAGC	2152968	
QY	460	TGGCGGCACACGTACGCTTGGATTAAGCGCGCGCTGCTGCTGGCCGATCAAGAGAAG	519	
Db	2152967	TGGCCGACACGCCAGCTTGGACAAGCGCGCGCGCTGCTGTTGGCGGTTCAGAAGAAG	2152908	
QY	520	TACGGCAACAAATACGCTGGGCGAGCTCATGATTTCTGGCTGGCAGCTGGCTTATGAG	579	
Db	2152907	TACGGCAAGAGCTCTATGGCGGACCTGATTTTTCGCGGCACTGCGCGCTTGGAA	2152848	
QY	580	TCCATGGGCTTACCTGCTTACGGCTTCTCTTTTCGGCGCGCTCGATATTTGGGAACCGAA	639	
Db	2152847	TCGATGGGCTTCAAGACGTTTCGGGTTTCGGCTTCGGCGCGGTCGACCACTGGGAGCCG	2152790	
QY	640	AAAGATATCTACTGGGGTGACGAAAAGAGTGCTGGCACCCTTCTGACGNACGCTACGGC	699	
Db	2152789	-ATGAGGTCTATTGGGGCAAGGAAGCACCTGGCTCG-----GCGATGAGCGTTACAGC	2152737	
QY	700	GACGTGAACAAAGCCAGAGACCATGGAAACCCCGCTGGCGGCTGCTCAAAATGGGCTGATC	759	
Db	2152736	GGTAACGGGATC-----TGGAGAACCCCGCTGGCGCGGTGCAGATGGGGCTTGATC	2152686	
QY	760	TATGTGAACCCGGAAGTGTTAAACGGCCACCCCTGATCCCGCTGAGAACCCGACAGCAGTA	819	

Db 2152685 TACGTGAACCGGAGGGCCGAACGCGAACCCCGACCCCATGGCCGGCGGTGCAGATT 2152626
Qy 820 CTTGAACCTTTGGCCCGGTATGGCGATGACAGCAGCAAAACCCGAGCCCTCACAGCTGGC 879
Db 2152625 GCGAGAGCTTTCCGGCGATGGCCATGAACAGCTCGAAACAGCGGGCTGATGTCGCG 2152566
Qy 880 GGCACACCGTCGGTAAATTGCAGGTAAATGGCAATGCCCTTGGTTAGCCCTTGACCCA 939
Db 2152565 GGTCAACATTTTCGGTAAGACCCATGGCGCGCGCGCGCGATGGTGGCCCGCGAACC 2152506
Qy 940 AAAGCCCTCTGACGTTGAACACAGGGCTTAGGTTGGGCGCAACCCCAACATGCAGGGCAAG 999
Db 2152505 GAGGCTGCTCGCTGGACGAGATGGCTTGGCTTGGAGAGCTGCTATGGCAACCGGAAC 2152446
Qy 1000 GCAAGCAACGCGGTGACCTCGGGGTATCGAAGGTGCTTGGACCAACCAACCCACGAAATTC 1059
Db 2152445 GGTAAAGACGCGATCAGCAGCGGATCGAGTCTGATGGAGCAACCCCGACGAATGG 2152386
Qy 1060 GATATGGGCTATTTCCGACCTGCTCTCGGCTACAAATTTGGGAACGAAAAAGAGTCCCTGCC 1119
Db 2152385 GACAACAGTTTCCCTCGAGATCCTGTACGGCTTACGAGTGGGAGCTGACGAAGAGCCCTGCT 2152326
Qy 1120 GGTGCCACCATTTGGGAACCGATTGACATCAAAAAGGAACAAAGCCGGTTGACGCCAGC 1179
Db 2152325 GGCCTTTGGCAATACACCGCAAGGACGGCGCCGCTCCCGCACCATCCCGACCCGCTTC 2152266
Qy 1180 GACCCCTCTATTTCGGCAACACCCGATCATGACCATGACGATGCGGATATGGGATAAAGTAAAT 1239
Db 2152265 GCGGGGCTC---AGGGCGCTCCCGACGATGTGGCCACTGACCTCTGCTCGGGTGGAT 2152209
Qy 1240 CCGACCTATCGCGCTATCTCGGAAATTCATGCGCGATCTGAGTACTTCAAGAAACT 1299
Db 2152208 CCGATCTATGAGCGGATCACCGCTGCTGGCTTGAACACCCCGAGGAATTTGGCGACGAG 2152149
Qy 1300 TTCGCAAGGGTGGTTCAAGCTGACGCACCGTGAACCTGGCGCGGAAATCACTATCATC 1359
Db 2152148 TTCGCAAGGGCTGTTCAAGCTGATCCACGACATGGTCCGCTTGGAGATACCTT 2152089
Qy 1360 GGGCCGGAAGTGGCGGAGAGACCTGATTTGGCAAGACCCGATTCGGGAGGTAAACAC 1419
Db 2152088 GGGCGCTGGTTCGCAAGCAGACCTGCTGTGGAGGATCCGCTCCCTGCGGTGAGCCAC 2152029
Qy 1420 GACTACTGCAAGAAGTGG-----TCAAGCAGAAATTTGCACAAAGTGGCCTG 1467
Db 2152028 GACCTGTGGGAAGCGGAGATTGCGAGCCCTTAAGAGCCAGATCCGGGCAATCGGGATTG 2151969
Qy 1468 AGCATTTAGTAGGTCTCCACCGCTTGGACAGTGCCTGCTATTCGCGGTTCGCAT 1527
Db 2151968 ACTGTCTACAGCTAGTTTCAACCGCATGGCGGGCGCTCGTCTGTTCCGTGTAGCGAC 2151909
Qy 1528 ATGCGCGCGGTGTAAAGGTGCGCGCATTCGTTGGCGCCCAACAGACGAGTGGCAGGGC 1587
Db 2151908 AAGCGCGCGCGCAACGGTGTGCGATCCGCCCTGCAAGCAGCAAGTGGGGTGGAGGTC 2151849
Qy 1588 AACGAGCCGGA---CGCCCTGGCAAGTGTGACGCTACGAGCTCTACGACAGATCT----- 1636
Db 2151848 AACGACCCCGAGCGGGATCTCGGAAGGTCTATTCGACCCCTTGAAGAGATCCAGAGTCA 2151789
Qy 1637 -----CTGCGCAGACCCGCGCTAGCATCGCGGAGCTGATGTTCTTGCGC 1680
Db 2151788 TTCAACTCCCGCGCGCGGGGAACATCAAAATGTCTTCGCGGACCTGCTGCTCGGT 2151729
Qy 1691 GGTAGCTAGCATCGGAAGACCGGAAAGCAGAGGTTACGATGTGCGGTTCCCTTC 1740
Db 2151728 GGCTGTGCGCGCATAGAAAGCAGCAAAAGCGGCTGGCCCAACATCACGGTGGCCCTTC 2151669
Qy 1741 CTGAAGGCGCTGCGCATGCGACCCGCGAGATGACCGACGAGACTCCTTCGCAACCGCTG 1800
Db 2151668 ACCCGGGCGCGAGTGGCTCGCAGGAACAAACCGACGTGGAAATCCTTTTSCGTGGCTG 2151609
Qy 1801 GAGCGCTGCGCGATGGCTTCGCAACTGCGCAGAAAGAGTATGTGTCAAGCCGGA 1860
Db 2151608 GAGCCCAAGCAGATGGCTTCGCAAACTTACCTCGGAAGGCAACCCCTTCCGCGCCGAG 2151549

Qy 1861 GAGATGCTGTGGATCGTGGCAGCTGATGGCTTTAAACCGCCCGGAAATGACCGTGGT 1920
Db 2151548 TACATGCTGCTCGCAAGCGCAACTGCTTACGCTCAGTCCCTGAGATGACGCTGGT 2151489
Qy 1921 CTGGCGGTATCCGCGTACTGGGCACCAACTATGGTGGCACAACACGCGGTATTAC 1980
Db 2151488 GTAGGTGGCTTGGCGTCTCGGGCAAACTACAAAGCGCTTACCGCTGGCGGTGTTAC 2151429
Qy 1981 GATTGTGAAGCCGATTTGACCAAGCACTTTTTTGTGAACCTGACCGATATGGGGAACAG 2040
Db 2151428 GAGCCCTCCGAGTCACTGACCAAGCACTTCTTGTGAACCTGCTCGCATGGGTATAC 2151369
Qy 2041 TGAAGC---CGGTAGGTAGCAACCGCTTACGAATTCGCGACCCCAAGACCGGTGCCGT 2097
Db 2151368 TGGAGCGCTCGCCAGCAGATGACGGGACCTACCAGGCAAGGATGGCAGTGGCAAGTG 2151309
Qy 2098 AAGTGGACCGCTCGCGGTGGATCTGGTATTTGGTTTCCAACTCGCTACTCGCTCTTAC 2157
Db 2151308 AAGTGGACCGCGACCGCTGGACCTGGTCTTTCGGGTCCAACCTCGGAGTTTCGGCGCTT 2151249
Qy 2158 GCAGAAAGTGTACGCCACGACGATAACGGCGAGAAGTTTCGTACAGAGACTTCGTGCGCC 2217
Db 2151248 GTCCAGGTCTATGGCGCGATGACGGCAGCGAGGTTCTGTGAGAGACTTCTGCTGCTGCC 2151189
Qy 2218 TGGACCAAGTGTGAACCGCGACCGTTCGACGT 2252
Db 2151188 TGGACAAGGTGATGAACCTCGACAGGTTCGACGT 2151154

RESULT 10
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 27.3%; Score 618.6; DB 3; Length 4411529;
Best Local Similarity 58.7%; Pred. No. 1.8e-162;
Matches 1253; Conservative 0; Mismatches 819; Indels 63; Gaps 8;
Qy 160 AACAAAGATTGGTGGCGCGGAAGGGTTGAACCTGGATATTTTGCATCAGCAAGATCGCAA 219
Db 2156007 AACAGGACTGGTGGCCCAACCGGCTCAATCTGAAGGTACTTGACCAAAACCCGCCCTC 2155948
Qy 220 TCAGACCGGATGGATTCGGGATTTCAACTTACCCTGGAAGATGACCAAGTACCAAGCTCGATTTCGAC 279
Db 2155947 GCTGACCGATGGGTGGCGCTTCGACTATCCGCGGAGGTTCGCGACCATCGACGCTTGAC 2155888
Qy - 280 GCGCTGAAGAAGATGTCACCGCTTGTATGACCATGACCAAGAGTGTGGCCCGCTGAC 339
Db 2155887 GCCCTGACCGCGACATCGAGAAGTGTATGCCAACCCTCGACGCTGGTGGCCCGCCGAC 2155828
Qy 340 TGGGGCACTACGCGGGTTTGTATGATCGTATGGCTTGGCACTCCGCTGGCACCTACCGT 399
Db 2155827 TACGGCCACTAGGCGCCCTGTTTATTCGGATGGCTGACGCTGCGCGCACCTTACCGC 2155768

QY 400 ATTGCTGATGGCGTGGGGGGTGTACCGGAAGCAGCGCTTTGCAACCGCTCAACTCC 459
Db 2155767 ATCCAGACGCGCGCGCGCGCGCGCGGATGACGAGCGTTCGGCGCGCTTAACAGC 2155708
QY 460 TGGCGGACAACTGACGCTGATGAAGCGCGCGCTGCTGCGCGGATCAAGAAG 519
Db 2155707 TGGCGGACAACTGACGCTGATGAAGCGCGCGCGCTGCTGCGCGGATCAAGAAG 2155648
QY 520 TACGGCAACAAATACGCTGGGCGAGACCTGATGATTTCTGGCTGGCACCGTGGCTTATGAG 579
Db 2155647 TACGGCAAGAACTCTCATGGCGGACCTGATGTTTTCGCGGCACTCGCGCTGGAA 2155588
QY 580 TCCATGGCTTACTGCTTACGCGCTTCTCTTTTCGCGCGCTGATATTTGGGAACCGGAA 639
Db 2155587 TCGATGGCTTCAAGACTTTCGGCTTCGGCTTCGGCGGCTGACCACTGCGGAGCGCG-- 2155530
QY 640 AAGATATCTACTTGGGTGACGAAAGAGAGTGTGCTGCGACCTTCTGACGAAACCTACGCG 699
Db 2155529 -ATGAGGTCTATTGGGCAAGGAAGCCACTGCTCG- - - - -GCGATGAGCGTTACAGC 2155477
QY 700 GACGTGAACAGCAGACACATGGAACACCCGCTGGCGCTGCTCCAAATGGGTCTGATC 759
Db 2155476 GGTAAAGCGGATC- - - - -TGGAGAACCCGCTGGCGCGGTGCGAGATGGGCTGATC 2155426
QY 760 TATGTGAACCGGAAGGTGTAAAGCGCCACCTGTATCCGCTGAGAACCGCAGCAGGTA 819
Db 2155425 TACGTGAACCGGAGGGCGGAACCGCAACCCGACCCCATGCGCGCGGCTCGACATT 2155366
QY 820 CTTGAAACCTTCGCCGCTATGGCGATGAACGAGAAACCCGACGCCCTTCACAGCTGCG 879
Db 2155365 CGGAGACGTTTCGGCGCATGGCCATGACGAGCTGGAACACGCGGCGCTGATCGTGGC 2155306
QY 880 GGCACACCGCTCGGTAAATGTCAGGTAAATGGCAATGCCCTCTCGTTAGCCCTGACCCA 939
Db 2155305 GGTACACTTTCGTAAGACCCATGCGCGCGCGCGCGGCTGCTGCGSCCGCAACC 2155246
QY 940 AAGCCCTCTGACCTTGAACACAGGCTTAGTTGGGCAACCCACATGACAGGCAAG 999
Db 2155245 GAGGCTGCTCGCTGAGCAGATGGGCTTGGGCTGGAAGAGCTCGTATGCAACCGCAACC 2155186
QY 1000 GCAAGAACCCGCTGACCTCGGCTATCGAAGTGTGGAACACCAACCCACGAAATTC 1059
Db 2155185 GGTAAAGACCGATACCGCGCATCGAGTGTGATGACGACACCCGCGAATGG 2155126
QY 1060 GATATGGCTATTTCGACCTGCTGCTGCGCTACATTTGGGAATGGAAGAGTCTGCGC 1119
Db 2155125 GACAACAGTTTCTCGAGATCTCTGACGCTACGAGTGGGAGCTGACGAAGAGCCCTGCT 2155066
QY 1120 GGTGCCACCATTTGGNAACGATTGACATCCAAAGGNAACCAAGCGGTTGACGCCAGC 1179
Db 2155065 GCGCTTGGCAATACACCGCAAGGACGCGCGCGGTGCGCGACCATCCCGGACCGCTTC 2155006
QY 1180 GACCCCTCTATTCCGCAACCCGATCATGACCGATCGCGATATGGCGATTAAGGTAAAT 1239
Db 2155005 GCGGGGCC- - - AGGGCGCTCCCGCGAGATGCTGGCCACTGACCTCTCGCTGCGGGTGGAT 2154949
QY 1240 CCGACCTATCGGCTATCTCGGAAATAATTCATGCGCGATGCTGAGTACTTCAAGAAACT 1299
Db 2154948 CCGATCTATGAGCGGATACGCGCTGCTGCTGGAACACCCCGAGGAATTTGGCCGACGAG 2154889
QY 1300 TTGCGGAAGCGGTTCAGCTGACGACCGGTGACCTGGGCCCGAATACACTTACATC 1359
Db 2154888 TTGCGGAAGCGGTTCAGCTGACGACCGGTGACCTGGGCCCGAATACACTTACATC 2154829
QY 1360 GGCAGGAGTGGCGGACGAAGACCTGATTGGAAGACCCGATTTCCGCGAGTAAACACC 1419
Db 2154828 GGCAGGAGTGGCGGACGAAGACCTGATTGGAAGACCCGATTTCCGCGAGTAAACACC 2154769
QY 1420 GACTACTGCGAAGAGTGG- - - - -TCAAGCAAGAAATTTGCACAAAGTGGCCCT 1467
Db 2154768 GACCTGTCGCGCAAGCGAGATTGCGCAGCTTAAGAGCGAGATCCCGGCGCATCGGATTG 2154709
QY 1468 AGCATTAGTGAATGCTCCACCAGCTTGGGACAGTGGCCGCTACTTATCGCGGTTCCGAT 1527

Db 2154708 ACTGCTTCACAGCTAGTTTTCAGCCGATGGCGCGCGCTCGTCTGCTTCCGTGGTAGCGAC 2154649
QY 1528 ATGGCGCGCGTCTACAGCTGCGCGCATTTGCTTGGCCCCACAGAAGAGTGGCAGGCG 1587
Db 2154648 AAGCGCGCGCGCGCAACAGTGTGCTCGCATCCGCTGCAGCGCAAGTTCGGGTGGGAGGT 2154589
QY 1588 AAGGAGCCGA- - -GCGCTTGGGAAAGTGTGAGCGTCTACGAGCAGATCT- - - - - 1636
Db 2154588 AAGGACCCCGACGGGATCTTGGCAAGTCTATTCGCAACCTTGAAGAGATCCAGGAGTCA 2154529
QY 1637 - - - - -CTGCCGACACGGCGCTAGCATCGCGACGCTTACGATGCGGAGTTCGCTTCC 1680
Db 2154528 TTCAACTCCGCGCGCGCGGGAACATCAAAAGTGTCTTTCGCCGACCTCGCTCGCTCGGT 2154469
QY 1681 GGTAGGTGATGGCATCGAAGAAAGCGGCAAGAGTGTACGATGCGGAGTTCAGTGTGCGGCTTCCCTTC 1740
Db 2154468 GGTGTGCCGCCATAGAAAGCAGCAAGGCGGCTGGCCACAACTACACGGTGGCGCTTC 2154409
QY 1741 CTGAAAGGCGGTGGCGATGCGACGCGCGGAGATGACCGACGACGACTCTTCGCAACCGCTG 1800
Db 2154408 ACCCGGGCGCGACGGATGCTCGCAGGAACAACCGACGCTTGAATCTTTGCGCTGCTG 2154349
QY 1801 GACCGCTGCGCGATGCTTCCGCAACTGCGCAAGAAAGAGTATGTGTTGAAGCGCGAA 1860
Db 2154348 GAGCCCAAGCAGATGCTTCCGAAACTACCTCGGAAGGCAACCGGTTGCGGCGCGAG 2154289
QY 1861 GAGATGCTGTGATGCTGCGGACGCTGATGGGCTTAACCGCGCGGAAATGACCGTGTGCTG 1920
Db 2154288 TACATGCTGCTGCAAGGCGAACCCTGTACGCTCAGTGCCTTGAGATGACGGTGTGCTG 2154229
QY 1921 CTGGGCGGTATGGCGCTACTGGGCAACAACTATGTTGGCAACCAACGCGCTATTACACC 1980
Db 2154228 GTAGTGGCGCTGCGCTCTCTCGCGCAAACTACAAGCGCTTACCGCTGGCGGTGTTACCC 2154169
QY 1981 GATTGTGAAGGCGAGTTGACCAACGACTTTTTTGTGAACCTGACCGATATGCGGGAACAGC 2040
Db 2154168 GAGGCTTCCGAGTCACTGACCAACGACTTCTTCTGTGAACCTGCTGACATGSGTATCACCC 2154109
QY 2041 TGAAGC- - -CGGTAGGTAGCAACGCTACGAAATCCGCGACCGCAAGCGGTGCGGTG 2097
Db 2154108 TGGAGCGCTCGCCAGCAGATGACGGGACCTACCGGCAAGGATGGCAGTGGCAAGGTG 2154049
QY 2098 AAGTGAACCGCTTCGGGGTGGATCTGTTATTTGGTTCCAACTCGCTACTGCGCTCTTAC 2157
Db 2154048 AAGTGAACCGCGACGCGCTGGACCTGCTTCCGGGTCCAACCTCGGAGTTGCGGCGCTT 2153989
QY 2158 GCAGAAAGTGTACGCCGAGGACGATACGGCGAGAGTTCGTCAGAGACTTCGTCGCGGCC 2217
Db 2153988 GTCGAGGTCTATGGCGCGATGACGCGAGCCGGAAGTTGCTGCGAGACTTCGCTGCTGCC 2153929
QY 2218 TGGACCAAGTATGAACCGCGCGCTTTTCGAGT 2252
Db 2153928 TGGGACAAAGTATGAACCTCGACAGTTCGAGT 2153894

RESULT 11

US-08-418-782-1
; Sequence 1, Application US/08418782
; Patent No. 5658733
; GENERAL INFORMATION:
; APPLICANT: Cockerill, Franklin R.
; APPLICANT: Kline, Bruce C.
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Isoniazid Resistant Strains
; TITLE OF INVENTION: of M. Tuberculosis
; NUMBER OF SEQUENCE: 22
; CORRESPONDENCE ADDRESS:
; ADDRESS: Schwegman, Lundberg & Woessner
; STREET: 3500 IDS Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA

ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/418,782
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Woessner, Warren D.
REGISTRATION NUMBER: 30,440
REFERENCE/DOCKET NUMBER: 150.141US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-339-0331
TELEFAX: 612-339-3061
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2235 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-418-782-1

Query Match 26.8%; Score 606.6; DB 1; Length 2235;

Best Local Similarity 58.6%; Pred. No. 1.1e-160;

Matches 1252; Conservative 0; Mismatches 819; Indels 64; Gaps 9;

```
QY 160 AACAAAGATTGGTGGCGGAGGTTGAACCTGGATATTTTGGCATCAGCAAGATCGGAAA 219
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 112 AACAGAGCTGGTGGCGCAACCGGCTCAATCTGAAGTACTGACCAAAAACCGCGCGTC 171

QY 220 TCAGACCCGATGGATCGCGATTTCAACTACCGTCAAGAAGTACCAAGCTCGATTTCGAC 279
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172 GCTGACCCGATGGTGGCGGCTTCGACTATGCCCGGAGTTCGGACCATCGAGTTGAC 231

QY 280 CGCGTGAAGAAAGATGTCACGCGCTGATGACCGATAGCCAAAGTGGTGGCGCGCTGAC 339
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 232 GCCCTGACGGGGACATCGAGGAAGTATGACCACTCGGACCTCGGACCGTGGTGGCGCGGAC 291

QY 340 TGGGGGCACTACCGCGGTTTGATGATCGGTATGGCTTGGCACTCCGCTGGCACCTACCGT 399
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 292 TACGGCCACTACGGGCGCTGTTATCCGGATGCGTGGCACGCTGGCGGCACTACCGC 351

QY 400 ATTGCTGATGCGGTGGGGCGGTGTACCGGAAGCCAGCGCTTTGACACCGCTCAACTCC 459
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 352 ATCCACGACGCGCGCGGCGCGCGGGGCGGCGATGACGGTTGCGCGCGCTTAACAGC 411

QY 460 TGGCCGACACGCTCAGCGCTGGATAAAGCGCGCGCTGCTGTGTGGCCGATCAAGAAGAAG 519
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 412 TGGCCGACACGCGAGCTTGGACAAGCGCGCGCGGCTGTGTGGCGGTCAGAAGAAG 471

QY 520 TACGGCAACAAATCAGCTGGGAGACCTGATGATTCCTGGCTGGCACCGGTGCTATGAG 579
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 472 TACGGCAAGAAGCTCTCATGGCGGACCTGATTTGTTTCGCGGCAACTGCGCGCTGGAA 531

QY 580 TCCATTGGCTTACCTGCTTACGGCTTCTTTTCGGCGCGCTGCGATATTTTGGGAACCGGAA 639
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 532 TCGATGGGCTTCAAGAGCTTCGCGGTTTCGGCTTGGCGGGGTGCGACAGTGGGAGCCCG -- 589

QY 640 AAGATATCTACTGGGTGACGAAAAGAGTGGCTGGCACCTTCTGACGAACGCTAGCGC 699
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 590 -ATGAGGTTATTGGGCAAGGAAGCAACCACTGGCTCG - - - - -CGGATGACGTTACAGC 642

QY 700 GACGTGAACAGCAGAGACCATGGAACCCGCTGGGGGTGTCRAAATGGGTCTGATC 759
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 643 GGTAAAGGGGATC - - - - -TGAAGAACCCGCTGGCGCGGTGTCAGATGGGCTGATC 693

QY 760 TATGTGAACCGGAAGTGTAAACGGCACCTGATCCGCTGAGACAGCAGCAGGTA 819
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 694 TACGTGAACCGGAGGGGCCGACGGCAACCCGACCCCATGTCGGCGGGCGGTGACAT 753
```

```
QY 820 CTTGAAACCTTTCGCCCGTATGGCGATGAACGACGAAAAAACCCGAGCCCTCACAGCTGGC 879
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 754 CGCGAGACGTTTCGGCGCATGGCCATGAACGACGTCGAAACAGCGCGCTGATCGTCGCG 813

QY 880 GGCACACCGCTCGGTAATTGTACAGGTAATGGGCAATGCCTCTGCTGTTAGCCCTGACCCA 939
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 814 GGTACACATTTTCGTAAGACCCATGGCGCGCGCGCGCGATCTGGTCGGCGCCGAAACCC 873

QY 940 AAAGCTCTGACGTTGAAACACAGGCTTAGTTGGGCAACCCCAACATCATCAGCGCAAG 999
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 874 GAGCTCTCGCTGGAGCAGATGGGCTTGGCTTGGAGAGCTCGTATGGCACCGGAAACC 933

QY 1000 GCAAGCAACCGCTGACCTCGGATCGAAGGTCTTGGACCAACCAACCCCAAGAAATTC 1059
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 934 GGTAAAGACGCGATCACGACGCGCATCGAGGTCGTATGGAGCAACACCCCGACGAATGG 993

QY 1060 GATATGGCTATTTTCGACCTGCTGTTGGCTACAAATTTGGGACATGAAAAAGAGTCTGCC 1119
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 994 GACAAAGATTTCTCTGAGATCTCTGACGGCTACAGTGGGAGTGAAGAGCCCTGCT 1053

QY 1120 GGTGCCACCATTTGGAAACCGATTGACATCAAAAAGGAACAAAGCCGTTGAGCCGACG 1179
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1054 GGCCTTGGCAATACACGCGCAAGGACGGCGCGCGTGGCGCACCATCCCGGACCCGCTTC 1113

QY 1180 GACCCCTTATTCGCCCAACCCCGATCATGACCGATCGGATATGGGATATAAGGTAAT 1239
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1114 GCGCGGC - - - AGGCGCTCCCGCAGCATGCTGGCCACTGACCTCTCGCTCGGCTGGAT 1170

QY 1240 CCGACCTATCGCGCTNTCTGGGAAAAATTCATGGCCGATCTGTAGTACTTCAAGAAACT 1299
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1171 CCGATCTAGAGCGGATCAGCGCTCGCTGGCTGGAACACACCCCGAGGAATTCGGCGACGAG 1230

QY 1300 TTCGGAAGGCGTGGTTCAAGCTGACGACCGCTGACCTGGCGCCCGAAATACGTTACATC 1359
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1231 TTCGCCAAGGCTGGTACAGCTGATCCACCGAGACATGGTCCGCTTGCAGATACCTT 1290

QY 1360 GGC CGGAGAGTGGCGGACAGAGACTGATTTGGCAAGACCCGATTCGGGCAAGTAACACC 1419
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1291 GGGCGGCTGGTCCCAAGCAGACCGCTGCTGTGGAGGATCCGCTCCCTCGCGTCAAGCCAC 1350

QY 1420 GACTACTGCGAAGAGTGG - - - - -TCAAGCAGAAATTTGCACAAAGTGGCGTG 1467
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1351 GACCTCTCGCGGAAGCGGAGATTGCCAGCGCTTAAGAGCCAGATCCGGGCAATCGGGATG 1410

QY 1468 AGCATTAGTAGAGTGTCTCCACCGCTTGGACAGTGCCTACTTATCGCGGTTCGAT 1527
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1411 ACTGCTCACAGCTAGTTTCGACCGCATGGCGGGCGGCTGCTGTTCCGTGGTAGGCAC 1470

QY 1528 ATGCGCGCGGTGTAAAGGTGCCCGCATTCGTTGGCCCCACAGACAGTAGTGGCAGGGC 1587
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1471 AAGCGCGCGCGCAACGGTGGTCCGATCCGCTGACGCCACCAAGTCCGGTGGGAGGTC 1530

QY 1588 AACGAGCGGA - - - GCGGCTGGCAAGTGTGAGGCTTACGAGCAGATCT - - - - - 1636
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1531 AACGACCCCGAGGGGATCTCGCAAGGTCTATTCGACCTCTGGAAGAGATCAGAGATCA 1590

QY 1637 - - - - -CTGCCGACACCGCGCTAGCATCGCGGAGCTGATGCTGTTGCGC 1680
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1591 TTCACCTCCGGCGCGCGGGGAAACATCAAAAGTGTCTTCGCGGACCTCGTGTGCTCGGT 1650

QY 1681 GGTAGCCTTAGCTCGAAGAAAGCGGAAAGCAGCAGGTTACGATGTGCGGTTCCCTTC 1740
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1651 GCGTGTGCCGCAATAGAAAGCAGCAAGCGGCTGGCCACACACATCAGCTGCGCTTC 1710

QY 1741 CTGAAGAGCCGTGGCGATGCGACCGCGGAGATGACCGAGCAGACTCCTTCGACCGCTG 1800
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1711 ACCCGGGCGCGACGGATGCGTCGAGGAACAAACCGACGTGAATCTTTTCCGCTGCTG 1770

QY 1801 GAGCGCTGGCGGATGGCTTCGCAACTGGCAGAAAGAGAGTATGTGTGAACCGCGAA 1860
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1771 GAGCCCAAGGCGAGATGGCTTCCGAAACTTACCTCGGAAAGGGGCAACCCCTTCCGCGCGAG 1830
```



```
Db 1114 GGGGGCC--AGGGGCTCCCGAGGATGCTGGCCACTGACCTCTCGCTCGGGTGGAT 1170
QY 1240 CCGACCTATCGCGCTAATCGGAAATAATCATGCGCGATCCCTGAGTACTTCAAGAAACT 1299
Db 1171 CCGATCTATGAGCGGATACGGGTCTGGCTGGAACACCCCGAGGAATGGCCGACGAG 1230
QY 1300 TTGCGCAAGCGTGGTTCAGCTGACGACCGCTGACCTGGGCGCGGAATCACTTACATC 1359
Db 1231 TTGCGCAAGCGTGGTTCAGCTGATCCACCGAGATGGTCCCGTTCGGAGATACCTT 1290
QY 1360 GGGCCGGAAGTCCCGCAGAAAGCTGATTTGGCAAGACCCGATTCGGCAGGTACAC 1419
Db 1291 GGGCCGCTGTCGCCAAGCAGACCTGCTGGCAGGATCCGCTCGGTGACCCAC 1350
QY 1420 GACTACTCGAAGAGTG--G-----TCAAGCAGAAATTCACAAAGTGGCGTG 1467
Db 1351 GACCTGTCGGGAGACCGAGATTCGACGCTTAAGAGCCAGATCCGGGCATCGGGATTG 1410
QY 1468 AGCATTTAGTATGATGCTCCACCGCTTGGACAGTGCCGTAATTCAGCGTTCGAT 1527
Db 1411 ACTGCTCACAGTATTTGACCGCATGGGCGGCGTGTGCTGCTGTAGCGAC 1470
QY 1528 ATGCGGGGGTGTCTAACGTCGCCGCTTGGCTTGGCCCGCAGAACAGTGCGAGGGC 1587
Db 1471 AAGCGCGGCGGCCAACGGTGTGCTGCTCCGCTGACGACACAGTCCGGTGGAGGTC 1530
QY 1588 AAGGAGCCGA---GGGCTCGGGAAGTGTGAGCGTCTACGAGCAGATCT----- 1636
Db 1531 AAGCGCCCGGAGGGATCTGGCAAGTCAATTCGACCGCTGGAAGAGATCCAGGAGTCA 1590
QY 1637 -----CTCCGACACCGCGGCTAGCATCGCGACGTGATCTCTGCGC 1680
Db 1591 TTCAACTCGCGGCGCGGGGAACATCAAGTGTCTTCGCGACCTCTGCTGCTGCT 1650
QY 1681 GGTAGCGTAGCATCGAAGAGCGCGAAGCAGAGTTACGATGTGCGCGTTCCTCTC 1740
Db 1651 GGTGTGCGCCATAGAGAAGCAGCAAGCGGCTGGCCACACATCAGTGCCTCTC 1710
QY 1741 CTGAAGGCGGTGGGATCGGACCGCGGAGATGACGCGCAGACTCTTCGACCGCTG 1800
Db 1711 ACCCGGGCGCAGCATGCTGCGAGGAACAAACGAGTGGATCTTTGCGGTGCTG 1770
QY 1801 GAGCCCTGCGGATGCTTCCGCAACTGCGCAGAGAAGAGATGATGTGGAAGCGGAA 1860
Db 1771 GAGCCCAAGCAGATGCTTCGGAACCTACCTCGGAAGGGCAACCCGTTGCGGCGGAG 1830
QY 1861 GAGATGCTGCTGATCGTGGCAGCTGTAGGGCTTAACCGGCGCGGAAATGACCGTGTG 1920
Db 1831 TACATGCTGCTGACAAAGGGAACCTGCTTACGCTCAGTGCCTGAGATGACGTGCTG 1890
QY 1921 CTGGGCGGTATGGCGTACTGGGACCAACTATGTTGGCAGCAACACACGCGGTATCAC 1980
Db 1891 GTAGTGGGCTGCGGCTCTCGG-GCAAACTACAAGCGCTTACCGGTGGCGGTTCACC 1949
QY 1981 GATTGTGAAGGCCAGTTGACCAACGACTTTTGTGAACCTGACCGATATGGGAACAGC 2040
Db 1950 GAGGCTCGAGTCACTGACCAACGACTTCTGTTGAACCTGCTGACATGGGTATCAC 2009
QY 2041 TGAAGC---CGTAGGTAGCAACGCTACGAAATCCGCGACCGCAAGACCGGTGCGGTG 2097
Db 2010 TGGGACCTTCGCGACAGATGACGGGACCTACGAGGCAAGGATGCGCAGTGGCAAGGTG 2069
QY 2098 AGTGACCGCTTCGCGGTGATCTGGTATTTGGTTTCAACTGCGTACTGCGCTTTAC 2157
Db 2070 AAGTGACCGGACGCGGTGGACCTTGTCTTGGGTTCCAACTGCGAGTTGGGGGCTT 2129
QY 2158 GCAGAGTGTACGCCAGCAGATACGCGGAGAGTTGCTCAGAGACTTCGTCGCGGCC 2217
Db 2130 GTCGAGGTCTATGGCGCCGATGACGCGCAGCGAGGTTGCTGAGGACTTCGTCGTGTC 2189
QY 2218 TGGACAAAGTGTATGAACGCGACCGCTTTCGAGCT 2252
Db 2190 TGGACAAAGTGTATGAACCTCGACAGCTTCGAGCT 2224
```

```
RESULT 13
US-08-852-219-1
; Sequence 1, Application US/08852219
; Patent No. 5922575
; GENERAL INFORMATION:
; APPLICANT: Cockerill, Franklin R.
; APPLICANT: Kline, Bruce C.
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Isoniazid Resistant Strains
; TITLE OF INVENTION: of M. Tuberculosis
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Muehling, Raasch & Gebhardt, P.A.
; STREET: 119 No. 5922575th Fourth Street, Ste. 203
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,219
; FILING DATE: 07-May-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sandberg, Victoria A.
; REGISTRATION NUMBER: 41,287
; REFERENCE/DOCKET NUMBER: 230.00010130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1226
; TELEFAX: 612-305-1228
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2235 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-852-219-1
```

```
Query Match 26.8%; Score 606.6; DB 2; Length 2235;
Best Local Similarity 58.6%; Pred. No. 1.1e-160;
Matches 1252; Conservative 0; Mismatches 819; Indels 64; Gaps 9;
QY 160 AACAAAGATTGGTGGCGGAAGGGTTGAACCTGGATATTTTGATCAGCAAGATCGCAA 219
Db 112 AACAGGACTGGTGGCCCAACCGGCTCAATCTGAAGGTACTGCACCAAAACCGCGCTC 171
QY 220 TCAGACCCCATGGATCCGGATTTCAACTACCGTGAAGAGTACGCAAGCTCGATTCGAC 279
Db 172 GCTGACCCCATGGTGGCGGTTGACATATGCGCGGAGGTGCGGACCATCGACGTTGAC 231
QY 280 GCGCTGAAGAAGATGCTCCACGCGTTGATGACCATAGCAGAGTGGTGGCCCGCTGAC 339
Db 232 GCGCTGACCGGGACATCGAGGAAGTATGACCACTCGCAGCCGTGGTGGCCCGCGAC 291
QY 340 TGGGGGCACTACGGCGGTTTGTATGATGATCCGTATGGCTTGGACATTCGCTGGCAGCTACCGT 399
Db 292 TAGCGCACTACGGCGGCTGTTTATCCGGATGGCGTGGCAGCTGCGGSCACTACCGC 351
QY 400 ATTGCTGATGGCGGTGGGGCGGTGTACCGGAGCAGCGCTTTTCACCGCTCACTCC 459
Db 352 ATCCAGCAGCGCGCGCGCGCGCGCGCGCGCATCGAGGGTTTCGCGCGCTTAACAGC 411
QY 460 TGGCCCGACAACCTGACGCTGGATAAAGCGCGCGCTGCTGTGTGGCGGATCAAGAAGAG 519
Db 412 TGGCCCGACAACCGCAGCTTGGACAAGCGCGCGCGCTGCTGTGTGGCGGTTCAAGAAGAG 471
```


[illegible]

Db 1226 AGTTCGGCAAGCGCTGGTACAAGCTGATCCACCGAGACATGGTCCCGTTGCGAGATACC 1285
QY 1358 TCGGCCCGGAAGTCCGCGCAGAAAGACCTGATTTGGCAAGACCCGATTCGCGCAGGTAAACA 1417
Db 1286 TTGGGCGCGTGGTCCCAAGCAGACCCCTGCTGGCAGGATCCGGTCCCTGCGGTCAAGCA 1345
QY 1418 C--CGACTACTCGGAGAGTGGTCAAGCAGAAATTCACAAAGTGGCCTGAGCATTT-- 1473
Db 1346 CGACTTCGTCGGCGAAGCAGATTGCCAGCCTTTAAGAGCCAGATCCGGGCAATCGGGATTGA 1405
QY 1474 -----AGTCAGATGGTCTCACCGCTTGGACAGTGGCCGCTACTTATCCGGGTTCGGATA 1528
Db 1406 CTGTCTCAGAGCTAGTTTCAGCGGATGGCGCGCGCTGCTGCTCCGTGGTAGCGACA 1465
QY 1529 TCGCGCGCGGTGCTTAACGGTGGCCGATTCGTTGGCCGCCACAGAACAGTGGCAGGGCA 1588
Db 1466 AGCGGGCGCGGCCAACGGTGGTCATCCGCTGCAGCCACAAAGTGGGGTGGAGGTCA 1525
QY 1589 ACSAGCCGAGCGCCTGGCGAAGTCTGAGCGTCTACGAGCAGATCTCTGCCG----- 1642
Db 1526 ACAGCCCGACGGATCTGGCAAGGTCAATTCGCACCCCTGAAGAGATCCAGGAGTCATTCA 1585
QY 1643 -----ACACCGCGCTAGCATCGCGACGCTGATCGTTCTGGCGGTAGCGTAG 1690
Db 1586 CTCGGCGCGGAAACATCAAGTGTCTTCGCCGACCTGCTGCTCGGTGGCTGTGCGC 1645
QY 1691 GCATCGAGAAAGCCGGAAGCAGCAGGTTAGCGATTGCGCGTTCCTTCCTCAAGGCC 1750
Db 1646 CACTAGAGAAAGCAGCAAGCGGCTGGCCACAACATCAGGTGCCCTTACCCCGGGCC 1705
QY 1751 GTGGCGATCGGACCGCGAGATGACCGGACGAGACTCTTCGACCGCTGGAGCCGCTGG 1810
Db 1706 CGCACGATCGTCGCGAAGCAACCGAGCTGGAATCTTTGCCGCTGCTGGAGCCCAAGG 1765
QY 1811 CGATGGCTTCGCAACTGGCAGAGAAAGATGATGTGTGAAGCCGGAAGAGATGCTGC 1870
Db 1766 CAGATGGCTTCGAAACTACCTCGGAAAGGGCAACCGTTGCCGCGCGAGTACATCGCTGC 1825
QY 1871 TGGATCGTCGCGAGCTGTAGTGGCTTAACCGCGCCGGAATGACCGCTGCTGCTGGCGGTA 1930
Db 1826 TCGACAAGCGCAACTGTCTACGCTCAGTGCCTCGATGACGGTGTGTTAGGTGGCC 1885
QY 1931 TGGCGGTACTGGCACCACCTATGGTGGCACCAACACCGCGGTATTCACCGATTTGAAG 1990
Db 1886 TGGCGTCTCGCGCAAACTACAAGCGCTTACCCTGGGCGCTGTTACCGAGGCGCTCG 1945
QY 1991 GCCAGTTGACCAACGACTTTTGTGAACCTGACCGATATGGGGAACAGCTGGAAGC--- 2047
Db 1946 AGTCACTGACCAACGACTTCTCGTGAACCTGCTCGACATGGGTATCACCTGGGAGCCCT 2005
QY 2048 CGGTAGGTAGCAACCGCTACGAAATCCGCGACCGCAAGACCGGTGCCGTGAAGTGCACCG 2107
Db 2006 CGCCACGATGACGGGACCTACCGAGGCAAGGATGGCAGTGGCAAGGTGAAGTGCACCG 2065
QY 2108 CCTCGGGGTGATCTGGTATTTGGTTCCAACTCGCTACTGCGCTCTTACGCGAAGTGT 2167
Db 2066 GCAGCCGCTGGACCTGGTCTTGGGTCCAACTCGGAGTTGGGGCGCTTGTGAGGTCT 2125
QY 2168 ACSGCCAGGAGGATACGGCGAAGTCTGTACAGACTTCGTGCGCGCTGACCAAAAG 2227
Db 2126 ATGCGCGGATGACGCGGAGGGAAGTCTGTACAGGATTCGTGCTGGTGGGACAAGG 2185
QY 2228 TGATGAACCGCCACCGTTTCGACGT 2252
Db 2186 TGATGAACCTCGACAGTTCGACGT 2210

Search completed: October 8, 2003, 15:33:54
Job time : 143.677 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 7, 2003, 19:31:42 ; Search time 16.4119 Seconds
(without alignments)
2134.727 Million cell updates/sec

Title: US-09-884-889-8
Perfect score: 4002
Sequence: 1 MENHKSSSYNTNGKC.....VKDFVKAQKVDLDRFDLK 745

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2826	70.6	740	1	CATB_STRRE
2	2716	67.9	740	1	CATB_STRCO
3	2714.5	67.8	735	1	CATB_BACST
4	2691	67.2	737	1	CATB_CAUCR
5	2584	64.1	749	1	CATB_LEGPN
6	2544	63.6	740	1	CATB_MYCBO
7	2541.5	63.5	746	1	CATB_MYCTU
8	2485.5	62.1	753	1	CAT2_NEUCR
9	2447	61.1	739	1	CATB_MYCSM
10	2444.5	61.0	726	1	CATB_SALTI
11	2442.5	60.6	726	1	CATB_SALTY
12	2423.5	60.6	726	1	CATB_ECOLI
13	2421.5	60.5	730	1	CATB_HALMA
14	2401.5	60.0	720	1	CATB_HALNL
15	2388	59.7	752	1	CAT1_MYCFO
16	2380.5	59.5	737	1	CATB_YERPE
17	2265	56.6	741	1	CATB_ARCFU
18	2265	56.6	741	1	CATB_ARHOC
19	2139	31.0	576	1	CCPR_YEAST
20	316	7.9	361	1	APX1_ARATH
21	244	6.1	249	1	APX1_PEA
22	229	5.7	249	1	APU_THETU
23	136	3.4	1861	1	NARG_BACSU
24	122.5	3.1	1228	1	NARG_RHIME
25	120.5	3.0	973	1	PER4_ARATH
26	119.5	3.0	315	1	GAG_BAEVM
27	119.5	3.0	537	1	TRC4_ECOLI
28	119.5	3.0	1861	1	TRC4_ECOLI
29	117.5	2.9	1279	1	APU_THESA
30	116.5	2.9	783	1	YGJK_ECOLI
31	115.5	2.8	787	1	XPKA_LACPE
32	113.5	2.8	882	1	ODP1_PSEAE
33	113	2.8	2561	1	PPS1_BACSU

34 113 2.8 3321 1 PCN2_HUMAN
35 112.5 2.8 1444 1 ADP1_MYGE
36 111.5 2.8 797 1 SYM_ARATH
37 111 2.8 1173 1 DP3A_PSEFL
38 109.5 2.7 807 1 AFSK_STRGR
39 109 2.7 1246 1 NARG_ECOLI
40 108 2.7 784 1 OSTA_SALTI
41 107 2.7 797 1 AF32_HUMAN
42 107 2.7 910 1 SYL_NEIMA
43 106.5 2.7 746 1 GUN1_STRRE
44 106.5 2.7 1616 1 SLAP_BACCI
45 106 2.6 292 1 EFTS_MYCPE

ALIGNMENTS

RESULT 1
CATB_STRRE STANDARD; PRT: 740 AA.
AC 087864;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Peroxidase/catalase (EC 1.11.1.6) (Catalase-peroxidase).
GN CPFB.
OS Streptomyces reticuli.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1926;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tu45;
RX MEDLINE=99231840; PubMed=10217488;
RA Zou P., Borovok I., Ortiz de Orue Lucana D., Muller D., Schrempf H.;
RT The mycelium associated Streptomyces reticuli catalase-peroxidase,
RL Microbiology 145:549-559(1999).
CC -! FUNCTION: BIFUNCTIONAL, EXHIBITING BOTH A CATALASE AND
CC BROAD-SPECTRUM PEROXIDASE ACTIVITIES (BY SIMILARITY).
CC -! CATALYTIC ACTIVITY: 2 H(2)O(2) - O(2) + 2 H(2)O.
CC -! COFACTOR: Binds 2 protoheme IX and 2 iron ions per tetramer (By
CC similarity).
CC -! SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL
CC PEROXIDASE/CATALASE SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Y14317; CAAT4698.1; -.
DR HSSP; P00431; ICCA.
DR InterPro; IPR000763; Bac_ctase/prase.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00141; peroxidase; 1.
DR PRINTS; PR00458; PEROXIDASE.
DR TIGRfam; TIGR00198; cat_per_HPI; 1.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS00436; PEROXIDASE_2; 1.
DR PROSITE; PS50873; PEROXIDASE_4; 1.
KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.
FT ACT_SITE 105 105 BY SIMILARITY.
FT ACT_SITE 109 109 BY SIMILARITY.
FT METAL 272 272 IRON (HEME AXIAL LIGAND).
SQ SEQUENCE 740 AA; 81345 MW; E21860AFE4B4A40E CRC64;

Query Match 70.6%; Score 2826; DB 1; Length 740;
Best Local Similarity 71.4%; Pred. No. 9.4e-177;
Matches 531; Conservative 77; Mismatches 128; Indels 8; Gaps 6;

Db 360 --APELIIPDAHPAKKHPRLTTDLRLDPIYGPISRRFYENPEEFADAFARAWFKLT 417

QY 422 HRDMGPKVYLGPEVPQEDLIWQDIPDVSHPLVDENDIEGLKAKILESGLTVSELSTA 481

Db 418 HRDMGPKSLYLGPEVPEETLIWQDPLPEPEGEVDAEDVATLTKLLESGLSLVQVTTA 477

QY 482 WASASTFRNSDKRGANGARILAPQKDWENNPPOOLARVLKTEGIEDBNQAOQSDNKA 541

Db 478 WASASTFRNSDKRGANGARILPEQGMVEVNEPDELAQVLRVLEGVQREFNSG-SGAK 536

QY 542 VSLADLIIVLAGCAGVEKAKDAGHEVQVFPNPGRADATAEOTDVAFALEPAADCFRNY 601

Db 537 VSLADLIIVLGSAAEKAEKAGFPVFPFAAGRVDAETEDTAESFALEPTADCFRNY 596

QY 602 IKPEHKVSAEMLVDRAQLLSAPEMTALVGMHVLGTNYDGSQHGVTINKPGOLSNDF 661

Db 597 LKGNRLPAEFLLLDRANLLTILSAPEMTALVGLRVLAGHQOQSLGVFTTRPGSLTND 656

QY 662 FVNLDDLTAKWRASDESOKVPEGRDFTKGEVKSCTRVLDLIFGNSSELRAEAUYGCADS 721

Db 657 FVNLDDLTAKWRASDESOKVPEGRDFTKGEVKSCTRVLDLIFGNSSELRAEAUYGCADS 716

QY 722 EKFVKDFVAKAWKMDLDRDL 744

Db 717 GEKFVDFVAAWKVMDLDRDL 739

RESULT 3

CATA_BACST STANDARD; PRT; 735 AA.

ID CATA_BACST STANDARD; PRT; 735 AA.

AC P14412.

AT 01-JAN-1990 (Rel. 13, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Peroxidase/catalase (EC 1.11.1.6) (Catalase-peroxidase).

GN PERA OR CAT.

OS Bacillus stearothermophilus.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.

OX NCBI_TaxId=1422;

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC STRAIN-ATCC 8005 / JAM11001.

RX MEDLINE-9911698; PubMed-2670897;

RA Loprasert S., Negro S., Okada H.;

RT "Cloning, nucleotide sequence, and expression in *Escherichia coli* of the *Bacillus stearothermophilus* peroxidase gene (perA).";

RL J. Bacteriol. 171:4871-4875(1989).

RN [2]

RP REVISIONS.

RA Trakulaleamsai S., Aihara S., Miyai K., Suga Y., Yomo T., Negro S.,

RA Urabe I.;

RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE-9911698; PubMed-9920270;

RA Matsura T., Miyai K., Trakulaleamsai S., Yomo T., Shima Y.,

RA Miki S., Yamamoto K., Urabe I.;

RT "Evolutionary molecular engineering by random elongation mutagenesis.;"

RL Nat. Biotechnol. 17:58-61(1998).

CC -1- FUNCTION: BIFUNCTIONAL, EXHIBITING BOTH A CATALASE AND BROAD-SPECTRUM PEROXIDASE ACTIVITIES.

CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.

CC -1- COFACTOR: Binds 2 protoheme IX and 2 iron ions per tetramer.

CC -1- SUBUNIT: Homodimer.

CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL PEROXIDASE/CATALASE SUBFAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).

CC

DR EMBL; M29876; AAA22655.1; -.

DR EMBL; AB020234; BAA37114.1; -.

DR PIR; J50520; J50520.

DR HSSP; P00431; IBJ9.

DR InterPro; IPR000763; Bac_ctase/prase.

DR InterPro; IPR002016; Peroxidase.

DR Pfam; PF00141; peroxidase; 1.

DR TIGRFams; TIGR00198; cat_per_HPI; 1.

DR PROSITE; PS00435; PEROXIDASE_1; 1.

DR PROSITE; PS00436; PEROXIDASE_2; 1.

DR PROSITE; PS00873; PEROXIDASE_4; 1.

KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.

FT ACT_SITE 97 97

FT ACT_SITE 101 101 BY SIMILARITY.

FT METAL 264 264 IRON (HEME AXIAL LIGAND).

SQ SEQUENCE 735 AA; 82989 MW; 7131204AA4BFABEF1 CRC64;

Query Match 67.8%; Score 2714.5; DB 1; Length 735;

Best Local Similarity 68.7%; Pred No 1.7e-169;

Matches 511; Conservative 75; Mismatches 139; Indels 19; Gaps 6;

QY 1 MENHKSGSYNTNTGKCPFTGSLQKSGAGGTGKRDWPNMNLNLGILRQHSLSDPN 60

Db 1 MENQ-----NRQNAACCPFGSVTQSS-NRTTKDWFNQLNLSILHQDRKTNPH 51

QY 61 DPDFYAEFKLDLAANKKDLAALMTDSQDWPADYGHYGPFFIRMAHWSAGTVRIGD 120

Db 52 DEEFYAEFKLDLAANKKDLAALMTDSQDWPADYGHYGPFFIRMAHWSAGTVRIGD 111

QY 121 RGGGSGSORFAPLNSWPDNANLDRKLLWPKYGRKISWADLMLTGNVALETMGF 180

Db 112 RGGASTGTQRFAPLNSWPDNANLDRKLLWPKYGRKISWADLMLTGNVALETMGF 171

QY 181 KTFGAGGADYVPEEDVYWGAEFTWLGDKRYEGDRELENPLGAVQMGLIYVPEGPN 240

Db 172 KTIGFGGRVDVWHPEDVYWGSEKWLASRYSGDRELENPLAAVQMGLIYVPEGPD 231

QY 241 KDPITAAARDIRETFGRMANDEEVVALTAGHTGKTHGAADAERYVGRPAAGIEM 300

Db 232 KDPKAAANDIRETFRRMGMDDEEVVALTAGHTGKTHGAAGPA-THVGPPEAAPIEAQ 290

QY 301 SLGWNTYGTGAGADTITSGLEGANTKTPTQNSNFFNLFGYEWELTSPAGAYQWPK 360

Db 291 GLGWLSYKKGKSGSDTITSGIEGANTPTPTQNDTSYFDMLFYDHWLTSFAGAWQWAV 350

QY 361 DGAGAGTIPDAHDPKSHAPFMTTDLALRMDPDYKISRYYENPEDEPADAFARAWFKL 420

Db 351 DPDEKDLAPDAEDPSKVPMTMMTTDLALRFPDPEVEKJARRFHQNPPEEFAAFARAWFKL 410

QY 421 THRDMGPKVRYLGPVPEQEDLIWQDIPDVSHPLVDENDIEGLKAKILESGLTVSELST 480

Db 411 THRDMGPKTRYLGPVPEQEDLIWQDIPDPEVDYELT-EAEIEEKAKILNSGLTVSELST 469

QY 481 AWASASTFRNSDKRGANGARILAPQKDWENNPPOOLARVLKTEGIEDBNQAOQSDNK 540

Db 470 AWASASTFRNSDKRGANGARILAPQKDWENNPPOOLARVLKTEGIEDBNQAOQSDNK 523

QY 541 AVSLADLIIVLAGCAGVEKAKDAGHEVQVFPNPGRADATAEOTDVAFALEPAADCFRNY 600

Db 524 KYSIADLIIVLGSAAEKAEKAGDAGFDVYKVPFFGRGDATQEDTDESFAVLEPFDAGFRN 583

QY 601 YTKPEHKVSAEMLVDRAQLLSAPEMTALVGMHVLGTNYDGSQHGVTINKPGOLSN 660

Db 584 YQKQSVSPPEELLVDKALLGLTAPEMTALVGLRVLAGHQOQSLGVFTTRPGSLTND 643

QY 661 FVNLDDLTAKWRASDESOKVPEGRDFTKGEVKSCTRVLDLIFGNSSELRAEAUYGCAD 720

Db 644 FVNLDDMTYEWVPTDSG--IYEIRDKTKTEVTRATRVLDLIFGNSSELRAEAUYQAOD 701

FT	ACT_SITE	106	106	BY	SIMILARITY
FT	METAL	268	268	IRON (HEWE AXIAL LIGAND).	
FT	SEQUENCE	737 AA;	80165 MW;	D900742D04E02E97 CRC64;	
Query Match		67.2%;	Score 2691;	DB 1;	Length 737;
Best Local Similarity		69.5%;	Pred. No. 5.9e-168;		
Matches	508;	Conservative	64;	Mismatches	145;
				Indels	14;
				Gaps	6;
QY	15	NTGGKCPFTGGSLKQSGAGGKTNRDWWPNMLNGLIRHQSSLSDPNDPDPDYAEFPKLLD	74		
DB	18	NTIAGKCP-----MGHGRGPNRDNWPOSRLLEGLNHAPRNPMPGEAPDYAEAFKSLD	70		
QY	75	LAAVKKDLAALMTDSODWMPADYGHYGPFFIRNAWHSAGTYRIGDCRGCGGSSQSRFAPL	134		
DB	71	LDVAVSDLHALMTDSOEWWMPADYGHYGGFIRLAWHAAGTYRTTDRGGGAGGGQQRFAPL	130		
QY	135	NSWPDNANLDKARLLWPKQYGRKISWADLMILGNVALETMGFTKTFGAGGRADWVE	194		
DB	131	NSWPDNTNLDKARRLLWPKQYKAGKLSWADLVVGVNVALESNGFTKTFGAGGRADQWE	190		
QY	195	PEEDVTWGAETEWLGRKVEGDERELNPLGAVOMGLIYVNPPEGNGKPDPIAARDIRET	254		
DB	191	PEE-LYWGPESTWLDKRYSGERELDSPLGAVOMGLIYVNPPEGNGPDPLASARDIRET	249		
QY	255	FGRMANDRETVALIAGHTGCTHCAADAERYGVGEPAAAGTEEMSLGWKNTYGTGCHGA	314		
DB	250	FARMANDDETVALIAGHTGCTHCAAGAGDA-SLVGYVEGGGAIEAQFGWASKHGTTGKP	308		
QY	315	DTITSGLEGAWTKTPTQWNNPFFENLFGYEWELTKSPAGAYQWKPKDGAGAGTIPDAHPD	374		
DB	309	DAITGSGPEVIWQTPTTWSNHFENLFGYEWELTKSPAGAKQWQAKN--APADIPDAFPD	366		
QY	375	SKSHAPFMUTDLALRMDPDYKISRYRYENPDEFADAFAKAWYKLTTHRDMGPKVRYLGP	434		
DB	367	NKTHVPRMLTSLALRFDPAYEKISRREYENPDQFADAFARAWFKLTHRDMGPIGRYLGP	426		
QY	435	EYPOEDLIWQDPTDVSHPPLVDENLLEGLKAKLTESGLTVSELVSTAWASASTFRNSDKR	494		
DB	427	LYPKELIWDQDPIPAVDHPLADKDLAAUKAKLATGLSASDLVSTAWASASTYRQSDKR	486		
QY	495	GGANGARIRLAPQKQWEVNNPQQLARVLTKLEGIOEDFNQAQSDNKAVSLADLVLAGCA	554		
DB	487	GGANGARIRLAPQKQWAVNNPVLAKVLAALGVQKDFNAGAGGKKISLADLVLAGGAA	546		
QY	555	GVEKAAKAGHGVQVFPNPGRADATAEQTDVZAFALPAADGFRNYIKP-EHKYSAEEM	613		
DB	547	AIEKAAKDAGSTVTPVAPGRMDASAEQTDHAFISFEALERSDGFNRGPGKHYNAPAEA	606		
QY	614	LVDRAGLLSLAPENTALVGGMRVLGTYNGDSQGHVFTNKPQLSNSDFVNLDLNTKKR	673		
DB	607	LVDRAGLLSLGPELTVLVGGVLRVIGANNADGSKGVFTNRPALNSDFVNLDSMETTW-	665		
QY	674	ASDESCKVFEGRDFKTGTGKVGSTRTVDLIFGNSNELRALAEVYGADSEKVPKDFVKAW	733		
DB	666	-SPTAANAFAGHDKRSSEPRWTATRVDLIFGSHAELRAFAEYVACADSQEKFCVDFVTAW	724		
QY	734	AKVMOLDREFDL 744			
DB	725	NKVMNADRLDL 735			
RESULT 5					
CATA_LEGPN					
ID	CATA_LEGPN	STANDARD;	PRT;	749 AA.	
QY	Q9WXB9;				
AC	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	Peroxidase/catalase (EC 1.11.1.6) (Catalase-peroxidase).				
GN	KATG.				
OS	Legionella pneumophila.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales				
OC	Legionellaceae; Legionella				

Query Match 63.68; Score 2544; DB 1; Length 740;
 Best Local Similarity 64.38; Pred. No. 2.3e-138;
 Matches 480; Conservative 88; Mismatches 168; Indels 10; Gaps 8;

Qy 2 ENRHSSSTVNTWTKGKPTGSLKOSAGGKTKNDWPNMNLGLTLRSHSLSPND 61
 Db 3 EOHPIITETTGASNG-CPVVGHKYVVEGG--NODWPNRLNLKVLHONPAVADPMG 59
 Qy 62 PDFDYAEFKKDLAALMTDSQDMADYGHYGFPIRMMAHWSAGTYRIGDGR 121
 Db 60 AAFDYAAEVATIDVLTALTRDIEEVMTTQPMWADYGHYGFPIRMMAHWSAGTYRIGDGR 119
 Qy 122 GGGSSGQRAPLNSWPNANLDRKLLWPIKOKYGRKISWADMLTGNVALETGFK 181
 Db 120 GCAGGGHQRAPLNSWPNANLDRKLLWPIKOKYGRKISWADMLTGNVALETGFK 179
 Qy 182 TFGAGGADWPEEDVYGAETWGLDKRYEGDRELENPLGAVOMGLIYVNPENGGK 241
 Db 180 TFGGFGVQWDEDE-VYWGKEATWGLDERYSGRDLENPLAAYVOMGLIYVNPENGN 238
 Qy 242 PDPTAAARDIRETFGRAMDEETVALIAGHTFGKTHGAADAKEYVYGREPAAAGIEMS 301
 Db 239 PDMAAAVDIRETFGRAMDEETVALIAGHTFGKTHGAADAKEYVYGREPAAAGIEMS 297
 Qy 302 LGWKTYCTGHGADITLISLEGATKPTQWNSNFFENLGYEWELTKSPAGATQWPKD 361
 Db 298 LGWKSSYGTGKDAITSGIEVWNTPTKWDNSFLGYEWELTKSPAGATQWPKD 357
 Qy 362 GAGAGTIPDPAH-DPSKSHAPFMTDLALRMDPDYKISRYENPDEFADAFKANYKL 420
 Db 358 GAGAGTIPDPGGPGRS--PTMLATDLSLRVDPIYERITRWLEHPELADFAKANYKL 415
 Qy 421 THRMGPKRYLGPVEPOEDLIWODPIDVSHPLVDENLEGLAKILEGLTSELVST 480
 Db 416 IHRDMGPVARYLGPVLPKQTLWQDPVAVSHDLVGEAETASLQSLASLTYSQVST 475
 Qy 481 AWASAFRSDKRGKGANGARILAPQKDWENPNQ-QLARVLTLEGIQEDFNQASDN 539
 Db 476 AWAAAFRSDKRGKGANGARILAPQKDWENPNQ-QLARVLTLEGIQEDFNQASDN 535
 Qy 540 KAVSLADLVLAGCAGVEKAKADAGHEVQVFPNPGRAATAEQDVFADFALEPAADGFR 599
 Db 536 IKVSFADLVLAGCAGVEKAKADAGHEVQVFPNPGRAATAEQDVFADFALEPAADGFR 595
 Qy 600 NYKPEHKVSAEMLVDRALQSLASPEMTALVGMVRLGNTYDGSQHGVTNPKGOLSN 659
 Db 596 NYLKGKGNLPAEYMLDKANLLTSAPEMTVVLGGLRVLAGANYKRLPLGVFTSESLTN 655
 Qy 660 DFFVNLDLNTKRASDESKVFRGDKTGEVKNWGTVDLFGNSSELRLAALVYGA 719
 Db 656 DFFVNLDMGITWEPSPADGTYGKRD-GSGKRVKWTGSRVDLVFGNSSELRLAALVYGA 714
 Qy 720 DSEKFKVFKVAKAWYMDLDRFDLK 745
 Db 715 DAOPKFEVDFVAAWKVMDLDRFDV 740

RESULT 7

ID CAT1_MYCTU STANDARD; PRT: 740 AA.
 AC Q08129; Q05444; Q05046; Q05051; Q05052; Q05053; Q05054;
 AC Q05055; Q57215; Q57274;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 GN Peroxidase/catalase T (EC 1.11.1.6) (Catalase-peroxidase T).
 DE KATG OR RV1908C OR MT1959 OR MTCV180.10.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 ON NCBI_TaxID=1773;
 RX NCBI_TaxID=1773;
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.

RC STRAIN=H37RV;
 RX MEDLINE=93308108; PubMed=8320241;
 RA Heym B., Zhang Y., Poulet S., Young D., Cole S.T.;
 RT "Characterization of the katG gene encoding a catalase-peroxidase
 required for the isoniazid susceptibility of Mycobacterium
 tuberculosis.";
 RT J. Bacteriol. 175:4255-4259(1993).
 RL [2]
 RC REVISIONS.
 RC STRAIN=H37RV;
 RA Cole S.T.;
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RC SEQUENCE FROM N.A.
 RC STRAIN=ATCC 25618;
 RA Cockerill F.R., Uhl J.R., Temesgen Z., Zhang Y., Stockman L.,
 RA Roberts G.D., Williams D.L., Kline B.C.;
 RT "Rapid identification of a point mutation of the Mycobacterium
 tuberculosis catalase-peroxidase (katG) gene associated with
 isoniazid resistance.";
 RT Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
 RL [4]
 RC SEQUENCE FROM N.A.
 RC STRAIN=INH-resistant strains;
 RA Marttila H.J., Soini H., Huovinen P., Viljanen M.K.;
 RT "katG gene mutations in isoniazid-resistant Mycobacterium
 tuberculosis strains isolated from Finnish patients.";
 RT Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 RL [5]
 RC SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence.";
 RT Nature 393:537-544(1998).
 RL [6]
 RC SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains.";
 RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 RL [7]
 RC PARTIAL SEQUENCE FROM N.A., AND ISONIAZID RESISTANCE.
 RP
 RC STRAIN=H37RV;
 RX MEDLINE=92365822; PubMed=1501713;
 RA Zhang Y., Heym B., Allen B., Young D., Cole S.T.;
 RT "The catalase-peroxidase gene and isoniazid resistance of
 Mycobacterium tuberculosis.";
 RT Nature 358:591-593(1992).
 RL [8]
 RC SEQUENCE OF 1-94 FROM N.A.
 RC STRAIN=H37RV;
 RA Song J., Deretic V.;
 RT Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 RL
 CC -1- FUNCTION: BIFUNCTIONAL, EXHIBITING BOTH A CATALASE AND
 CC BROAD-SPECTRUM PEROXIDASE ACTIVITIES. MAY PLAY A ROLE IN THE
 CC INTRACELLULAR SURVIVAL OF MYCOBACTERIA.
 CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) -> O(2) + 2 H(2)O.
 CC -1- COFACTOR: Binds 2 prothene IX and 2 iron ions per tetramer.
 CC -1- SUBUNIT: Homodimer (Probable).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and/or commercial

```

Db      710 FVAAWKVMNLDREDL 725

RESULT 12
CATA_SALTY STANDARD; PRT; 726 AA.
AC P17750;
DT 01-AUG-1990 (Rel. 15, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Peroxidase/catalase HPI (EC 1.11.1.6) (Catalase-peroxidase)
DE (Hydroperoxidase I).
DE KATG OR STM4106.
GN Salmonella typhimurium.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2;
RX MEDLINE=91117169; PubMed=2277629;
RA Loewen P.C., Stauffer G.V.;
RA "Nucleotide sequence of katG of Salmonella typhimurium LT2 and
RT characterization of its product, hydroperoxidase I.";
RL Mol. Gen. Genet. 224:147-151(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
CC -1- FUNCTION: Bifunctional, exhibiting both a catalase and
CC -1- broad-spectrum peroxidase activities.
CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) - O(2) + 2 H(2)O.
CC -1- COFACTOR: Binds 2 protoheme IX and 2 iron ions per tetramer.
CC -1- SUBUNIT: Homotrimer.
CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL
CC PEROXIDASE/CATALASE SUBFAMILY.
CC -1- SIMILARITY: Contains 1 heme regulatory motif (HRM) repeat.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/annouce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X53001; CAA37187.1; -.
DR EMBL; AE008891; AAL22946.1; -.
DR PIR; S12039; CSEBHT.
DR HSSP; P48534; IAPX.
DR StyGene; SG10190; katG.
DR InterPro; IPR000763; Bac_catase/prase.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00141; peroxidase; 1.
DR PRINTS; PR00458; PEROXIDASE.
DR TIGRfams; TIGR00198; cat_per_HPI; 1.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS00436; PEROXIDASE_2; 1.
DR PROSITE; PS00873; PEROXIDASE_4; 1.
KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide;
KW Complete proteome.
FT ACT_SITE 102 102 BY SIMILARITY.
FT ACT_SITE 106 106 BY SIMILARITY.
FT METAL 267 267 IRON (HEME AXIAL LIGAND).

```

```

FT REPEAT 15 20 HRM.
FT CONFLICT 71 71 Y -> YY (IN REF. 1).
FT CONFLICT 136 136 N -> T (IN REF. 1).
FT CONFLICT 223 223 G -> D (IN REF. 1).
FT CONFLICT 228 228 N -> T (IN REF. 1).
FT CONFLICT 233 233 D -> N (IN REF. 1).
FT CONFLICT 275 275 A -> P (IN REF. 1).
FT CONFLICT 545 545 G -> R (IN REF. 1).
FT CONFLICT 549 549 S -> H (IN REF. 1).
FT CONFLICT 553 553 A -> P (IN REF. 1).
FT CONFLICT 561 561 Q -> H (IN REF. 1).
FT CONFLICT 634 634 R -> K (IN REF. 1).
SQ SEQUENCE 726 AA; 79656 MW; 7C4BA4439E9FFAB CRC64;

Query Match 61.0%; Score 2442.5; DB 1; Length 726;
Best Local Similarity 63.5%; Pred. No. 9.4e-152;
Matches 468; Conservative 84; Mismatches 166; Indels 19; Gaps 7;

Qy 11 TYNTNTGKCPFTGGSLKQSGAGGKTKNRDWWPNMNLGILRQHSLSLSDPNPDYAEFF 70
Db 7 THNTLTGKCPFHGGHDSAGAGTASRDWWPNQLRYDLLNQHNSRNLGEGEDFYRKEF 66
Qy 71 KKLDAAVKDLAALMTDSODWNPADYGHYGPFFIRMAHWSAGTYRIGDGRGGGSGSOR 130
Db 67 SKLDYSALKGDLKALLTDSQPNWPNADMGSYGLFIRMAHWSAGTYRISIDGRGGAGRQOR 126
Qy 131 FAPLNSWPDNANLKDARLLWPIKQYGRKISWADLMILTCNVALETMGFTTFFGAGRA 190
Db 127 FAPLNSWPDNVSLDKARLLWPIKQYGRKISWADLIFLAGNVALENSGRTFFGAGRE 186
Qy 191 DWPEPEDEVYGAETEMIGDKRYEGDRELENPLGAVOMGLIYVNPENGPDPDIAARD 250
Db 187 DWPEPDLDVNWGDEKAWLTTHRHPA--LAKAPLAGATEMLIYVNPENGPDHSGEPLSAAA 244
Qy 251 IRETFGMMNDRETVALLAGGTFGKTHGAADAKEYVYGPAPAAAGTEMSLGKNTYGT 310
Db 245 IRATFGMMNDRETVALLAGGTFGKTHGA--AASHVGADPEAPAEQAQGLGWASSTGS 303
Qy 311 GHGADTTTSGLEGATKTPTQWSNNFFENLFGYEMELTKSPAGAYQMKPKDGAGATIPD 370
Db 304 GVGADATITSGLEVVTQPTQWSNVFFENLFGYEMVQTRSPAGATQPEAVD--APDIIPD 361
Qy 371 AHDPKSHAPFMTDTDLALRMDPYEKISRYRYENPDDEFADAFKAWYKLTTHROMGPKVR 430
Db 362 PFDPKSKRKPTMLVTLDTLRFDEFEKISRRLNDPQAFNAFAFAFWKLTTHROMGPKAR 421
Qy 431 YLGEVPEQEDLIWODPIPDVSHPLVD--ENDIEGLKAKILSGLTVSLVSTANASATF 488
Db 422 YIGEVPEKEDLIWODPLP---QPLYQPTQEDIIINLKAIAASGLSISEWVSVAASATF 478
Qy 489 RNSDKRGANGARIRLAPQKDWYNNPQQLARVLKTLLEGIOEDFNQAQSDNKAYSLADLI 548
Db 479 RGGDKRGANGARLALAPQRDWDVN--AVALRVLPVLEEIQKTNKA-----SLADII 529
Qy 549 VLACGAGVEKAAKADAGHEVQVFPNPGRADATATBOTDVEAFEALEPAADGFNRYLKEHKV 608
Db 530 VLAVGVGTEQAAAAAGVSISYFPFAPGRVDAQQDQTDIEMFSLLEPIADGFNRYRLDVS 589
Qy 609 SAEMLYDRAQLLSILSAPEMTALVGGMRVLGTNTYDGSQHGVTNKPGLSNDFFNVLIDL 668
Db 590 TTESLLIDKAAQLTLTAPEMTVLVGGMRVLGTNTYDGSQHGVTNKPGLSNDFFNVLIDL 649
Qy 669 NTKWRASDESDFEGRDFTGKVGKSGTRVDLIFGNSNSELRALAEVYGCADSEKFKVKD 728
Db 650 RYEMKPTDDANELEFEGDRLTGKVGKYATRADLVFGNSVLRALAEVYACSDAEKFKVKD 709
Qy 729 FVKAWAKVMOLDREDLK 745
Db 710 FVAAWKVMNLDREDLQ 726

RESULT 13
CATA_ECOLI

```

ID CATA_ECOLI STANDARD; PRT; 726 AA.
AC FI3029;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Peroxidase/catalase HPI (EC 1.1.1.6) (Catalase-peroxidase)
DE (Hydroperoxidase 1).
DE KATG OR B3942.
GN Escherichia coli.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=88314956; PubMed=3045098;
RA Triggs-Raine B.L., Doble B.W., Mulvey M.R., Sorby P.A., Loewen P.C.;
RT "Nucleotide sequence of katG, encoding catalase HPI of Escherichia
RT coli.";
RL J. Bacteriol. 170:4415-4419(1988).
RN [2]
RN SEQUENCE OF 1-339 FROM N.A.
RP STRAIN=K12 / MG1655;
RC MEDLINE=93347969; PubMed=8346018;
RX Plunkett G. III, Burland V.D., Daniels D.L., Blattner F.R.;
RA "Analysis of the Escherichia coli genome. III. DNA sequence of the
RT region from 87.2 to 89.2 minutes.";
RL Nucleic Acids Res. 21:3391-3398(1993).
RN [3]
RN SEQUENCE OF 309-726 FROM N.A.
RP STRAIN=K12 / MG1655;
RC MEDLINE=94089392; PubMed=8265357;
RX Blattner F.R., Burland V.D., Plunkett G. III, Sofia H.J.,
RA Daniels D.L.;
RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the
RT region from 89.2 to 92.8 minutes.";
RL Nucleic Acids Res. 21:5408-5417(1993).
CC - FUNCTION: Bifunctional, exhibiting both a catalase and
CC broad-spectrum peroxidase activities.
CC - CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC - COFACTOR: Binds 2 protoheme IX and 2 iron ions per tetramer.
CC - SUBUNIT: Homotetramer.
CC - INDUCTION: By hydrogen peroxide.
CC - PTM: The N-terminus is blocked.
CC - SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL
CC PEROXIDASE/CATALASE SUBFAMILY.
CC - SIMILARITY: Contains 1 heme regulatory motif (HRM) repeat.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M21516; AAA24040.1; -;
DR EMBL; LI9201; AAB03074.1; -;
DR EMBL; U00006; AAC43048.1; -;
DR EMBL; AB000468; AAC76924.1; -;
DR PIR; A65201; CSECHP.
DR HSP; P00431; LCYP.
DR SWISS-2DPAGE; P13029; COLI.
DR EcoGene; EGI0511; katG.
DR InterPro; IPR000763; Bac_ctase/prase.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00141; peroxidase; 1.
DR PRINTS; PR00458; PEROXIDASE.
DR TIGRfams; TIGR00198; cat_per_HPI; 1.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS00436; PEROXIDASE_2; 1.
DR PROSITE; PS00873; PEROXIDASE_4; 1.
KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide;
KW Complete proteome.

FT ACT_SITE 102 102 BY SIMILARITY.
FT ACT_SITE 106 106 BY SIMILARITY.
FT METAL 267 267 IRON (HEME AXIAL LIGAND).
FT REPEAT 15 20 HRM.
FT CONFLICT 621 621 A -> G (IN REF. 1).
SQ SEQUENCE 726 AA; 80023 MW; 24D32EBED5DE9BD6 CRC64;
Query Match 60.6%; Score 2423.5; DB 1; Length 726;
Best Local Similarity 63.6%; Pred. No. 1.6e-150;
Matches 467; Conservative 78; Mismatches 172; Indels 17; Gaps 7;
QY 12 YNTTGGKCPPTGGSLKQAGGKTKRDWPNMNLGILRQHSLSDDPNDPFDYAEKPK 71
DB 8 HNTATGKCPFHGGHQDQAGAGTTTRDWWPNQLRVDLLNQHNSRNPGLGEDFDYKERS 67
QY 72 KIDLAARKDLAALMTDSODWHPADYGHVGPFFIRMAHWSAGTYRIGDGRGGGSSORF 131
DB 68 KLDYGLKKDLKALLTESQFWPADWGSYAGLFIKRWAGGTYRISIDGGRGAGRGQQR 127
QY 132 APLNSWPDNANLDKARLLWPVKQYGRKISWADLMILTCNVLETMGKTFGAGGRAD 191
DB 128 APLNSWPDVSLDKARRLLWPVKQYGRKISWADLMILTCNVLETMGKTFGAGGRAD 187
QY 192 VNEPEEDVTGAETEWLGDYKRYEGDRELENPLGAVOMGLIYVNPENKPKPIAARDI 251
DB 188 VNEPDLVNWGDWKAWLTHRHEA--LAKAPLGATEMGLIYVNPENKPKPIAARDI 245
QY 252 RETEGRMANDEETVALIAGGHTFGTKGAADAKEYVGRPAAGTEESLGNKNTYGTG 311
DB 246 RATFGNMGNDEETVALIAGGHTLKGTHGAGTSTN-VGPDPEAAPIEEGGLWASTYSG 304
QY 312 HGADITTSLEGAWTKTPTQWNNFFENLFGYEWELTKSPAGAYQWKPKDAGAGTIPDA 371
DB 305 VGADAITSGLEVVTQTPTQWNNFFENLFGYEWELTKSPAGAYQWKPKDAGAGTIPDA 362
QY 372 HPDSKSHAPFMTLTDLALRMDPDYKISRRYENDEPADAKAWYKLTNRDMGPKVRY 431
DB 363 FDPSSKRRKPTMLVTDLTLEDFEPEKISRRFLNDPQAFNEAFARAWFKLTHRDMPKSY 422
QY 432 LGPEVPOEDLIWQDPIPD-VSHPLVDNDIEGLKAKILESLTVSLVSTANASATFRN 490
DB 423 IGPEVPKEDLIWQDPLPQPIYP--TEQDIIDLKFAIDSGLSVSELVSVANASATFRG 480
QY 491 SKRGANGARIRLAPQKDMWYNNPQQLARVLTLEQIEDPQAOQDNKAYSLADLIYL 550
DB 481 GDKRGANGARLALMPQDMDVN--AAAVRALPVLEKIQESGKA-----SLADIIVL 531
QY 551 ACCAGVEKAAKAGHEVQVFPNPGADATAEQTDVEAFEALEPAADGFRNYIKPEKVSA 610
DB 532 AGVGVGEKAAKAGHEVQVFPNPGADATAEQTDVEAFEALEPAADGFRNYIKPEKVSA 591
QY 611 EEMLVDRALQLLSIAPEMTALYGGMRVLCTNYDQSGHGVTKNPKQLSNDPFPVLLDLNT 670
DB 592 ESLLIDKAOQLTLTAPETALVGGMRVLGANGFGSKNGVFTDRVGVLSNDFVNLDMRY 651
QY 671 KWRASDESDFEGRDFTKTEYKWSGTRVLDLIFGNSSELRLAELAYVGCADSEKFKVDV 730
DB 652 EMKATDESKELEFGRORETEYKWSGTRVLDLIFGNSSELRLAELAYVGCADSEKFKVDV 711
QY 731 KAWAKVMDLDRFDL 744
DB 712 AAWVKVMNLDLDRFDL 725

RESULT 14

CATA_HALMA

ID CATA_HALMA STANDARD; PRT; 730 AA.

AC O59651;

DT 16-OCT-2001 (Rel. 40, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Peroxidase/catalase (EC 1.1.1.6) (Catalase-peroxidase).

GN PERA.


```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE005159; AAG220931.1; -
CC DR EMBL; AF069761; AAC23534.1; -
CC DR PIR; T44562; T44562.
CC DR HSP; P00431; 1RYC.
CC DR InterPro; IPR000763; Bac_ctase/prase.
CC DR InterPro; IPR002016; Peroxidase.
CC DR Pfam; PF00141; peroxidase.1.
CC DR PRINTS; PR00458; PEROXIDASE.
CC DR TIGRFS; TIGR00198; cat_per_HPT.1.
CC DR PROSITE; PS00435; PEROXIDASE_1; 1.
CC DR PROSITE; PS00436; PEROXIDASE_2; 1.
CC DR PROSITE; PS00873; PEROXIDASE_4; 1.
CC DR Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide; Plasmid;
CC KW Complete proteome.
CC FT ACT_SITE 79 79 BY SIMILARITY.
CC FT ACT_SITE 83 83 BY SIMILARITY.
CC FT METAL 248 248 IRON (HEME AXIAL LIGAND).
CC FT SEQUENCE 720 AA; 80476 MW; FB88823BCD3CB2F9 CRC64;
CC -----
Query Match 60.0%; Score 2401.5; DB 1; Length 720;
Best Local Similarity 59.7%; Pred. No. 4.4e-149;
Matches 447; Conservative 104; Mismatches 165; Indels 33; Gaps 4;
QY 1 MENHKHS-GSSYNTNTGKCPFTGGSLKQSGAGGCTKNRDWPNMLNLGILRQHSLSLDP 59
Db 1 MENEIDNFGTS-----DWWPNQLDLLELDQNSQQVDP 32
QY 60 NDPDFDYAEFPKDLAAYKKDLAALMTDSQDWPADYGHYGPFFIRMAHWSAGTYRGD 119
Db 60 NDPEFDYAEFPKDLAAYKKDLAALMTDSQDWPADYGHYGPFFIRMAHWSAGTYRGD 119
QY 33 YGEDFYAEAFEDLOLAAYKKDLLEMTDSKDWPADYGHYGPLFIRMAHWSAGTYRTFD 92
Db 33 YGEDFYAEAFEDLOLAAYKKDLLEMTDSKDWPADYGHYGPLFIRMAHWSAGTYRTFD 92
QY 120 GRGGGGSGSRAPLNSWDNANLQKARLLWPIKQYGRKISWADLMLLTGNVALETWG 179
Db 120 GRGGGGSGSRAPLNSWDNANLQKARLLWPIKQYGRKISWADLMLLTGNVALETWG 179
QY 93 GRGGAAGGRQLPPVDSWPNVNLQKARLLWPIKQYGRKISWADLMLLTGNVALETWG 152
Db 93 GRGGAAGGRQLPPVDSWPNVNLQKARLLWPIKQYGRKISWADLMLLTGNVALETWG 152
QY 180 RTFTGAGGRADVWPEEDVYWGAEW---LGDKRYEGDRELEPLGAVOMGLIYVNP 236
Db 180 RTFTGAGGRADVWPEEDVYWGAEW---LGDKRYEGDRELEPLGAVOMGLIYVNP 236
QY 153 FETYGAGGRKDDYTPDEAVDVGPEDEWETTSDD-RFDADGSLKWLPLGNTVGLIYVNP 211
Db 153 FETYGAGGRKDDYTPDEAVDVGPEDEWETTSDD-RFDADGSLKWLPLGNTVGLIYVNP 211
QY 237 GNGKPDPTAARDTRFTFGRMANDEETVALIAGHTFGKTHGAADAKEYVGREPAAAG 296
Db 237 GNGKPDPTAARDTRFTFGRMANDEETVALIAGHTFGKTHGAADAKEYVGREPAAAG 296
QY 212 GNGEPDPLEGSGAKNIRSFGRKMANDEKTVALLIAGHTFGKTHGAADAKEYVGREPAAAP 271
Db 212 GNGEPDPLEGSGAKNIRSFGRKMANDEKTVALLIAGHTFGKTHGAADAKEYVGREPAAAP 271
QY 297 IEEMSLGHKNTYGTGHGADTITSGLEGATWKTPTQWSNFFENLFGYEWELTKSPAGAYQ 356
Db 297 IEEMSLGHKNTYGTGHGADTITSGLEGATWKTPTQWSNFFENLFGYEWELTKSPAGAYQ 356
QY 272 IEKQGLGWENEGEGKGPDPTITSGLEGPMNTPTQWMSYVDNLLEYEWEPEKGGGAWQ 331
Db 272 IEKQGLGWENEGEGKGPDPTITSGLEGPMNTPTQWMSYVDNLLEYEWEPEKGGGAWQ 331
QY 357 WKPKDGAGAGTTPDARDPSKSHAPMLTDLALRMDPDYKISRRYENPNPEFADAPAKA 416
Db 357 WKPKDGAGAGTTPDARDPSKSHAPMLTDLALRMDPDYKISRRYENPNPEFADAPAKA 416
QY 332 WTKKSGELNESAPGVQDPTDIEDVWMLTDLVALKDDPDYREVLETFQENPREFQGSFKA 391
Db 332 WTKKSGELNESAPGVQDPTDIEDVWMLTDLVALKDDPDYREVLETFQENPREFQGSFKA 391
QY 417 WYKLTNRDMPKRVYLGSEVPEQEDLIWQDPIPDVSHPLVNDENDIEGLKAKILESLTVSE 476
Db 417 WYKLTNRDMPKRVYLGSEVPEQEDLIWQDPIPDVSHPLVNDENDIEGLKAKILESLTVSE 476
QY 392 WYKLTNRDMPKRVYLGSEVPEQEDLIWQDPIPDVSHPLVNDENDIEGLKAKILESLTVSE 451
Db 392 WYKLTNRDMPKRVYLGSEVPEQEDLIWQDPIPDVSHPLVNDENDIEGLKAKILESLTVSE 451
QY 477 LVSTAWASASTFRNSDKRGKGANGARIRLAPQKDEHVNPNQOLARVLKTLGIGEDFNQAO 536
Db 477 LVSTAWASASTFRNSDKRGKGANGARIRLAPQKDEHVNPNQOLARVLKTLGIGEDFNQAO 536
QY 452 LVKWTAWASASTFRNSDKRGKGANGARIRLAPQKDEHVNPNQOLARVLKTLGIGEDFNQAO 511
Db 452 LVKWTAWASASTFRNSDKRGKGANGARIRLAPQKDEHVNPNQOLARVLKTLGIGEDFNQAO 511
QY 537 SDNKAVSLADLIVLAGACVGAADKAGHEVQVFNPNGRADATAEOTDVEAFEALEPAAD 596
Db 537 SDNKAVSLADLIVLAGACVGAADKAGHEVQVFNPNGRADATAEOTDVEAFEALEPAAD 596
QY 512 SDDMRVSLADLIVLGGNAIEGAADAGYVDVFEFGRTDPTFEQTDVSEFEALKPKAD 571
Db 512 SDDMRVSLADLIVLGGNAIEGAADAGYVDVFEFGRTDPTFEQTDVSEFEALKPKAD 571
QY 597 GFRNYIKPEHKVSAEMLVDRALQLLSAPEMTALVGGMRVLGTYDQSGRGVFTNPKPGQ 656
Db 597 GFRNYIKPEHKVSAEMLVDRALQLLSAPEMTALVGGMRVLGTYDQSGRGVFTNPKPGQ 656

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 7, 2003, 19:31:42 ; Search time 16.5881 Seconds
(without alignments)
2134.727 Million cell updates/sec

Title: US-09-884-889-6

Perfect score: 4058

Sequence: 1 MNNASADLHSLQRCRAF.....RDFVAATKYNADREFDVAS 753

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2489	61.3	735	1 CATA_BACST	P14412 bacillus st
2	2216.5	54.6	739	1 CATA_MYCSM	O59557 mycobacteri
3	2168.5	53.4	740	1 CATE_STRE	O87864 streptomyce
4	2162.5	53.3	737	1 CATA_CAUCR	O31066 caulobacter
5	2149.5	53.0	740	1 CATE_STROCO	O91799 streptomyce
6	2095	51.6	726	1 CATA_ECCLI	P13029 escherichia
7	2086	51.4	726	1 CATA_SALTY	P17750 salmonella
8	2083	51.3	726	1 CATA_SALTY	O82303 salmonella
9	2063.5	50.9	746	1 CATA_MYCIT	O04657 mycobacteri
10	2050	50.5	752	1 CATL_MYCFO	O08404 mycobacteri
11	2042.5	50.3	740	1 CATA_MYCTU	O08129 mycobacteri
12	2036.5	50.2	740	1 CATA_MYCBO	P46817 mycobacteri
13	2003	49.4	753	1 CATE_NEUCR	O81182 neurospora
14	2002	49.3	749	1 CATA_LEGPN	O942b9 legionella
15	1998	49.2	730	1 CATA_HALMA	O59651 halococcus
16	1930	47.6	720	1 CATA_HALMA	O73955 halobacteri
17	1922.5	47.4	737	1 CATA_YERPE	O946b0 yersinia pe
18	1901.5	46.9	741	1 CATA_ARCFU	O28050 archaeoglob
19	1642.5	40.5	576	1 CATA_RHOCA	P37743 rhodobacter
20	267	6.6	249	1 APXL_ARATH	O05431 arabidopsis
21	248.5	6.1	361	1 CCPR_YEAST	P00431 saccharomyc
22	245.5	6.0	249	1 APXL_PEA	P04334 pium sativ
23	121	3.0	857	1 SYL_BACST	O87190 vibrio para
24	118	2.9	358	1 PE12_ARATH	O96520 arabidopsis
25	117	2.9	345	1 TL29_LYCES	O946b6 lycopersico
26	113.5	2.8	372	1 LICG_TRAVE	P20013 tramatete ve
27	113.5	2.8	1424	1 NC03_HUMAN	O956g9 h nuclear r
28	112	2.8	321	1 PE28_ARATH	O95867 arabidopsis
29	111.5	2.7	485	1 VG14_BPMO2	O64207 mycobacteri
30	111.5	2.7	844	1 PAC_KLUCI	P07941 kluyvera ci
31	109	2.7	441	1 PUR2_METAC	O84466 methanosarc
32	108.5	2.7	874	1 SLAP_BACLI	P04952 bacillus il
33	108	2.7	326	1 PER6_ARATH	O48677 arabidopsis

ALIGNMENTS

RESULT 1

CATA_BACST	ID	CATA_BACST	STANDARD;	PRT;	735 AA.
AC	P14412;				
DT	01-JAN-1990 (Rel. 13, Created)				
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DT	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	Peroxidase/catalase (EC 1.11.1.6) (Catalase-peroxidase).				
GN	PERA OR CAT.				
OS	Bacillus stearothermophilus.				
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.				
OX	NCBI_TaxID=1422;				
RN	[1]				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RC	STRAIN-ATCC 8005 / IAM11001;				
RA	MEDLINE=89359121; PubMed=2670897;				
RX	Loprasert S., Negro S., Okada H.;				
RT	"Cloning, nucleotide sequence, and expression in Escherichia coli of the Bacillus stearothermophilus peroxidase gene (perA).";				
RL	J. Bacteriol. 171:4871-4875(1989).				
RN	[2]				
RP	REVISIONS.				
RA	Trakulnaleamsai S., Aihara S., Miyai K., Suga Y., Yomo T., Negro S., Urabe I.;				
RL	Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE=99116998; PubMed=9920270;				
RX	Matsuura T., Miyai K., Trakulnaleamsai S., Yomo T., Shima Y., Miki S., Yamamoto K., Urabe I.;				
RT	"Evolutionary molecular engineering by random elongation mutagenesis.";				
RL	Nat. Biotechnol. 17:58-61(1998).				
CC	- - FUNCTION: BIFUNCTIONAL, EXHIBITING BOTH A CATALASE AND BROAD-SPECTRUM PEROXIDASE ACTIVITIES.				
CC	- - CATALYTIC ACTIVITY: 2 H(2)O(2) - O(2) + 2 H(2)O.				
CC	- - COFACTOR: Binds 2 protoheme IX and 2 iron ions per tetramer.				
CC	- - SUBUNIT: Homodimer.				
CC	- - SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL PEROXIDASE/CATALASE SUBFAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	-----				
CC	EMBL: M29876; AAA22655.1; -.				
DR	EMBL: AB020234; BAA37114.1; -.				
DR	PIR: JS0520; JS0520.				
DR	HSSP: P00431; 1BJ9.				
DR	InterPro: IPR000763; Bac_ctase/prase.				
DR	InterPro: IPR002016; Peroxidase.				
DR	Pfam: PF00141; peroxidase; 1.				

Q9K7S8 bacillus ha
Q02934 clostridium
P75354 mycoplasma
P08083 escherichia
Q01401 oryza sativ
P16332 mus musculu
Q9JW39 neisseria m
Q919P0 arabidopsis
P06875 escherichia
Q9FX85 arabidopsis
Q43872 arabidopsis
Q23474 arabidopsis

34 106 2.6 806 1 SYL_BACHD
35 106 2.6 879 1 GUNI_CLOTH
36 105.5 2.6 910 1 DNJM_MYCPN
37 105 2.6 387 1 CEAN_ECCLI
38 105 2.6 820 1 GLGB_ORYSA
39 104.5 2.6 748 1 MUTA_MOUSE
40 104.5 2.6 910 1 SYL_NEIMA
41 104 2.6 339 1 PE29_ARATH
42 104 2.6 846 1 PAC_ECCLI
43 103.5 2.6 350 1 PE10_ARATH
44 103 2.5 317 1 PE64_ARATH
45 101.5 2.5 348 1 PE40_ARATH

16-OCT-2001 (Rel. 40, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
Peroxidase/catalase T (EC 1.11.1.6) (Catalase-peroxidase T).
KATG
Mycobacterium smegmatis.
Bacteria: Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium.
NCBI_TaxID=1772;
[1]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 700084 / mc(2)155;
RX MEDLINE=97023930; Pubmed=8870251;
RA Billman-Jacobe H., Sloan J., Coppel R.L.;
RT "Analysis of isoniazid-resistant transposon mutants of Mycobacterium
smegmatis.";
RL FEMS Microbiol. Lett. 144:47-52(1996).
[2]
SEQUENCE FROM N.A.
RC STRAIN=mc(2)1216;
RA Engler O., Telenti A.;
Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: BIFUNCTIONAL, EXHIBITING BOTH A CATALASE AND
BROAD-SPECTRUM PEROXIDASE ACTIVITIES. MAY PLAY A ROLE IN THE
INTRACELLULAR SURVIVAL OF MYCOBACTERIA.
CC -!- CATALYTIC ACTIVITY: 2 H(2)O(2) -> O(2) + 2 H(2)O.
CC -!- COFACTOR: Binds 2 prothème IX and 2 iron ions per tetramer.
CC -!- MISCELLANEOUS: DELETIONS OR DEFECTS IN KATG GENE CAUSE ISONIAZID
(INH) RESISTANCE.
CC -!- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL
PEROXIDASE/CATALASE SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL: X98718; CA667268.1; -
EMBL: U46844; AAC45275.1; -
HSP: P00431; IBEK.
DR InterPro: IPR000763; Bac_ctase/prase.
DR InterPro: IPR002016; Peroxidase.
DR Pfam: PF00141; peroxidase; 1.
DR PRINTS: PR00458; PEROXIDASE.
DR TIGRfams: TIGR00198; cat_per_HPI; 1.
DR PROSITE: PS00435; PEROXIDASE_1; 1.
DR PROSITE: PS00436; PEROXIDASE_2; 1.
DR PROSITE: PS50873; PEROXIDASE_4; 1.
KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.
FT ACT_SITE 110 110
FT ACT_SITE 114 114 BY SIMILARITY.
FT METAL 277 277 IRON (HEME AXIAL LIGAND).
FT CONFLICT 51 51 L -> M (IN REF. 2).
FT CONFLICT 214 214 D -> RT (IN REF. 2).
FT CONFLICT 374 378 PTRTA -> AHEDG (IN REF. 2).
SQ SEQUENCE 735 AA; 81131 MW; A15A754A826F8D4 CRC64;
Query Match 54.6%; Score 2216.5; DB 1; Length 739;
Best Local Similarity 57.1%; Pred. No. 1.2e-147;
Matches 425; Conservative 98; Mismatches 198; Indels 23; Gaps 8;
QY 26 PRHRAIRER-----AMSGKCPVWHG--GNTSTGTSKNOWPGLNDILHQDRKS 74
DB 2 PEDRPEDSPPIGEAQTADAPAGGCPAGFGRIKPPVAGGSRNDRWPNQLNKLTKLPDVI 61
QY 75 DPMDPDFNYREVRKLDKDVHALMTDSQEWNPADMGWGHYGGGLMIRMAHWSAGTYRI 134
DB 62 NPLDEDFYRSVONLDVADRIDIVEMTSDQMPADFGHGPGLFIRMAHAACTYRV 121
QY 135 ADGRGGGTGSGQRFAPLNSWPDNYSLDKARLLWPIKKYKNSISWADLMILAGTVAYES 194


```

Db 122 SDGRGAGAGMORFAPLNSWPDNLSKARRLLVKKYKNSWADLVVAGNVALD 181
QY 195 MGLPAYGFSFGVRIWEPKDIYWGDEKWLAPSDERYGDVKNKPTMENPLAAVOMGLTY 254
Db 182 MGFRTAGFAGGREDRWEPEEDVYWGPEQEWL--DKRY--TGERDLENPLAAVOMGLTY 236
QY 255 VNPGEVGNHPDLRTAQOVLETFARMANDEKTAALTAGGHTVGNCHNGNNSALADPK 314
Db 237 VNPEGPNPNPDQASADIRETFGRMANNDVETAAALIVGGHTFGTHGNDASLVGPPE 296
QY 315 ASDVENOGLGNPNMCKASNAVTSIGEGAWTNTPTKFDGMGYFDLLFCYNWELAKSPAG 374
Db 297 AAPLEVLGWRNPGTGVGKDAIYSGLEVHTWPTKWDNSFLILYGNWELTKSPAG 356
QY 375 AHWHPEDIKKENKVDASDPISIRNPIMTDADMAIKVNPYRAICERFMDPEYKKTIF 434
Db 357 ANQWPKDNGWANS-VPLPTRAKTHPSMLTSDALRVDPPIYEQITRRWLDHPLEAEF 415
QY 435 AKAWFKLTHRLDGPKSRYIGPVPAEDLIWQDPIAGNTDYCEEVVKQ--IAQSLGIS 492
Db 416 AKAWFKLTHRLDGPVTRVYLGPEVPKDTWLQDNIIPAGNDLSDEVAKLKLIADSLTVS 475
QY 493 ENVSTANDSARTYRGSDMRGGANGARIRLAPONOGNEPERLAKVLSYIEQISADTG-- 550
Db 476 QLVSTAWRAASTFRSLDGGANGRIRLQPOLGWEANEPEDELAQVVRKYIEIQKASGIN 535
QY 551 ASIADVIVLAGSVGIEKAAAGYDVRVPFLKRGDGAETAMDADSFAPLEPLADGFRNW 610
Db 536 VSFADLVVLGNGVVEKAAAGFDVTPFTPGRGDQTEEDVDVSFAYLEPKADGFRNY 595
QY 611 OKKEYVTPPEEMLDRAQLMGLTGPEMVLGGMRVLTGNYGGTKHGVTCEGQLTNDF 670
Db 596 LQKSGDLPAEFKLIIDRANLLGLSAPEMTTLVGLRLVDVNHGKTHGVLTDPKGLATDF 655
QY 671 FYNLDMGNSWK--PVGSNAYEIRDKTKGAVKWTASRVLDLVFGNSLLRSYAEVVAODN 728
Db 656 FYNLDMSTAWKPSPADGTGTYIGTRAYGSPKWTGTRVDLVFASNSQLRALAEVDAEDS 715
QY 729 GEKVFYDFVAWTKYMNADRFDA 752
Db 716 KEKFKDFVAWTKYMNADRFDA 739

RESULT 3
CATB_STRRE STANDARD; PRT; 740 AA.
AC O87864;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Peroxidase/catalase (EC 1.11.1.6) (Catalase-peroxidase).
GN CPEB.
OS Streptomyces reticulatus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycetaceae; Streptomycetaceae; Streptomyces.
OX NCBI_Taxid=1926;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Tu45;
RX MEDLINE=99231840; PubMed=10217488;
RA Zou P., Borovok I., Ortiz de Orue Lucana D., Muller D., Schrempf H.;
RT its gene and regulation by Furs.*;
RL Microbiology 145:549-559(1999).
CC -1- FUNCTION: BIFUNCTIONAL. EXHIBITING BOTH A CATALASE AND
CC BROAD-SPECTRUM PEROXIDASE ACTIVITIES (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) -> O(2) + 2 H(2)O.
CC -1- COFACTOR: Binds 2 prothome IX and 2 iron ions per tetramer (By
CC similarity).
CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL
CC PEROXIDASE/CATALASE SUBFAMILY.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y14317; CAA74698.1; -.
DR HSSP; P00431; ICCA.
DR InterPro; IPR000763; Bac_ctase/prase.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00141; peroxidase; 1.
DR PRINTS; PR00458; PEROXIDASE.
DR TIGRfam; TIGR00198; cat_per_HPI; 1.
DR PROSITE; PS00435; PEROXIDASE 1; 1.
DR PROSITE; PS00436; PEROXIDASE 2; 1.
DR PROSITE; PS50873; PEROXIDASE 4; 1.
KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.
FT ACT_SITE 105 105 BY SIMILARITY.
FT ACT_SITE 109 109 BY SIMILARITY.
FT METAL 272 272 IRON (HEME AXIAL LIGAND).
SQ SEQUENCE 740 AA; 81345 MW; E21860AFE48A40E CRC64;

Query Match 53.4%; Score 2168.5; DB 1; Length 740;
Best Local Similarity 55.3%; Pred. No. 2.9e-144;
Matches 415; Conservative 114; Mismatches 188; Indels 33; Gaps 11;

QY 24 VSPRRATRRAM---SKCPCVHGG--NTSTGTSKDWHPGCLNLDILHQDRKSDPM 78
Db 1 MTENDHDAIVTDAKSGSGGCPVAHRAHPTGGGNROWPERLNLKILKANPAVANPLD 60
QY 79 PDFNTREREVRKLDFAALKKDVHMLTDSQEMWPAWGHYGGIMIRMAHSACTYRIADGR 138
Db 61 EDFDFAEAFKALDLAARVDAIEVLTTSQDWPAFGNYGPLMIRMAHSACTYRISDGR 120
QY 139 GGGGTGSGRFPALNSWPNVSLDKARRLLWPKKYKNSIWADLMILAGTVAYESMGLP 198
Db 121 GGAGAGQORFAPLNSWPNNGNLDKARRLLWPKKYKNSIWADLLITGNVALETMGFK 180
QY 199 AYGFSGFRVDIWEPEKDIYWGDEKWLAPSDERY--GDVKNKPTMENPLAAVOMGLTY 257
Db 181 TFGGGRADVWEAEEDVYWGPEETWL--DDRRYTGDD---RELENPLGAVOMGLTY 234
QY 258 EGVNGHPDLRTAQOVLETFARMANDEKTAALTAGGHTVGNCHNGNNSALADPKASD 317
Db 235 EGPNGNPDPIAAARDIRETFRRMANDEETVALIAGHTFGTKHAGPADHVCADPEAAS 294
QY 318 VENQGLGNGPNMCKASNAVTSIGEGAWTNTPTKFDGMGYFDLLFCYNWELAKSPAGAH 377
Db 295 LEEQGLWRSTYGTGKADAITSGLEVHTWPTKWDNSFLILYGNWELTKSPAGAHQ 354
QY 378 WEPIDIKKENK--VDASDPSIRNPIMTDADMAIKVNPYRAICERFMDPEYKKTIF 434
Db 355 W-----VAKNAPEIIPDAHDPSKKHPRMLTDLRLFPDIYEPISRRYEHEPEEFADAF 409
QY 435 AKAWFKLTHRLDGPKSRYIGPVPAEDLIWQDPIAGNTDYCEE-----VVKQIAQSGLS 490
Db 410 ARAWYKLTDRMGPKSLYLGPEVPEETLWQDPLPEREGELIDDADIALTKLLESGLS 469
QY 491 ISEMYSTANDSARTYRGSDMRGGANGARIRLAPONOGNEPERLAKVL-----SVYEQTS 546
Db 470 VSQLVTTAWASASTFRASDKRGANGARIRLAPORGWEVNDPDQLAQVLRILTENVQEPN 529
QY 547 ADTGA--SIADIVIVLAGSVGIEKAAAGYDVRVPFLKRGDGAETAMDADSFAPLEPL 603
Db 530 ASSGAKKVSADLVILVGGAGVEKAAKEAGFEIQVPPFTGPRVDATDEEHTDVSFEALEPT 589
QY 604 ADGFRNWKQKEYVWPKPEEMLDRAQLMGLTGPEMVLGGMRVLTGNYGGTKHGVTDC 663
Db 590 ADGFRNVLGKGNRLPAEYLLLDKANLLNLSAEMTVLVGGLVGLGANHQOOLGVTFTTP 649
QY 664 GOLTNDFFVNLTDGNSMKNPKVGSN--AYEIRDKTKGAVKWTASRVLDLVFGNSLLRSYAE 721

```

Db 650 GVLNDFVLLDNGTTTGWKATSEDTTFEGRDATGEVKWAGSRADLVFGSSELRALAE 709
 QY 722 VYAADNGEKFKYRDFVAAWTKYNNADRFV 751
 Db 710 VYASDDAKKFKYRDFVAAWHKYMDARFDL 739

RESULT 4

CATA_CAUCR STANDARD; PRT; 737 AA.
 ID CATA_CAUCR STANDARD; PRT; 737 AA.
 AC Q21066; 2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Peroxidase/catalase (EC 1.11.1.6) (Catalase-peroxidase).
 GN KATG OR CC3043.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
 OC Caulobacteraceae; Caulobacter.
 OX NCBI_TaxID=155892;
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 19089 / CB15;
 RX MEDLINE=21173698; PubMed=11259647;
 RA Nielsen M.W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
 RA Utterback T.L., Iran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RT *Complete genome sequence of Caulobacter crescentus.*;
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 RN [2]
 RP SEQUENCE OF 1-494 FROM N.A.
 RC STRAIN-ATCC 19089 / CB15;
 RX MEDLINE=38012985; PubMed=9352936;
 RA Steinman H.M., Fareed F., Weinstein L.;
 RT *Catalase-peroxidase of Caulobacter crescentus: function and role in
 RT stationary-phase survival.*;
 RL J. Bacteriol. 179:6831-6836(1997).
 CC -1- FUNCTION: BIFUNCTIONAL, EXHIBITING BOTH A CATALASE AND
 CC BROAD-SPECTRUM PEROXIDASE ACTIVITIES.
 CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) - O(2) + 2 H(2)O.
 CC -1- COFACTOR: Binds 2 protoheme IX and 2 iron ions per tetramer (by
 CC similarity).
 CC -1- INDUCTION: By hydrogen peroxide.
 CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL
 CC PEROXIDASE/CATALASE SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AF005967; AA25005.1; -;
 DR EMBL; AF027168; AAC45850.1; -;
 DR PIR; A87626; A87626.
 DR PIR; T45480; T45480.
 DR HSSP; P00431; IRYC.
 DR TIGR; CC3043; -;
 DR InterPro; IPR000763; Bac_ctase/prase.
 DR InterPro; IPR002016; Peroxidase.
 DR Pfam; PF00141; peroxidase; 1.
 DR PRINTS; PR00458; PEROXIDASE.
 DR TIGRFAMS; TIGR00198; cat_per_HPI; 1.
 DR PROSITE; PS00435; PEROXIDASE_1; 1.
 DR PROSITE; PS00436; PEROXIDASE_2; 1.
 DR PROSITE; PS00873; PEROXIDASE_4; 1.

RESULT 5

CATA_STRCO

ID CATA_STRCO STANDARD; PRT; 740 AA.
 AC Q9RJH9;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Peroxidase/catalase (EC 1.11.1.6) (Catalase-peroxidase).
 GN CPB OR SC0560 OR SCF73.07C.

KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide;
 KW Complete proteome.
 FT ACT_SITE 102 BY SIMILARITY.
 FT ACT_SITE 106 BY SIMILARITY.
 FT METAL 268 IRON (HEME AXIAL LIGAND).
 SQ SEQUENCE 737 AA; 80165 MW; D900742D04E02E97 CRC64;
 Query Match 53.3%; Score 2162.5; DB 1; Length 737;
 Best Local Similarity 56.3%; Pred. No. 7.5e-144;
 Matches 416; Conservative 103; Mismatches 185; Indels 35; Gaps 10;
 QY 33 ERAMSGKCPVHGGNTSTGTSTNKDWPENGLNLDLHODRKSDPMDPDFNYREVRKLD 92
 Db 16 EDNIAGKCPMGHG---RGPANRDWNPQSLRLEGLNHAPSRNPMGEAFDYAEAFKSLDL 71
 QY 93 DALKKDVHALMTDSQEWNPADWGHYGLMIRMAHWSAGTYRIADRCGGCGGTGQSRFAPLN 152
 Db 72 DAVVSDLHALMTDSQEWNPADFGHYGLFIRLAWHAAGTYRITDGRGAGGGGQGRFAPLN 131
 QY 153 SWPDNVSLDKARLLNPKKKYGNKISWADLMILAGTVAVESMGLPAYGFSFGRVDIWEP 212
 Db 132 SWPDNTNLDKARLLNPKKYGAKLSWADLYVLVGNVALESMTGFTFGAGRADWEP 191
 QY 213 EKDIYWGDEKEWLAPSDERYGDVKNKPTMENPLAAVQMGLIYVNPGEVNGHPDPLRTAQ 272
 Db 192 E-ELYWGPESTWL--DDKRYSGERE---LDSPLGAVQMGLIYVNPGEVNGHPDPLASARD 245
 QY 273 VLTEFARMANDEKTAALTAGGTVGCHGNGNASALAPDKASDVENQGLHGNPNMQG 332
 Db 246 IRTFARMANDEETVALIAGGHTFGKAHAGDASLVGVEPGEAIEAGFGWASKHG 305
 QY 333 KASNAVTSIGEGAWTNTPTKFDGMYFDLLFGYNWELKSPAGAHWE---PIDIKENK 388
 Db 306 KGPDAITGGPEVIWOTPTTRNSNHFDFNLFKYEWELTQSPAGAKQKQKAPADIP--- 361
 QY 389 PVASDPSIRHNPIMTDADMAIKVNPYRAICEKFMADPEYFKKTFKAKWFKLTHRDLP 448
 Db 362 --DAFDPNKHVPRMLTSDIALRFPAYEKISRREYENPDQFADAFARWFKLTHRDMP 419
 QY 449 KSVYIGVEPAEDLIWQDPIPAQNTDYCEB---VVKOKIAQSGLSISEMVSTAWDSART 504
 Db 420 IGYLGLPLVPEKEELIWDPIPAVDHPDADKDKTAALKAKLATGLSASDLVSTAWASAST 479
 QY 505 YRSDMRGGANGARIRLAPQNEQWQNEPERLAKVLSVYEQISADTGA-----STADY 556
 Db 480 YRQSDKRGANGARIRLAPQKDWAVNPPVLAKVLAALLEGVQKDFENASAGGKKISLADL 539
 QY 557 IVLGSGVIGIEKAAKAGYDVRVPLKRGDATAEMTDADSFAPLEPLADGFRWQ--KKE 614
 Db 540 IVLGGAALIEKAAKADAGTSVTPFPAPGRMDASAEQTDHAHSFEALPRSDGFRNYRGPGR 599
 QY 615 YVYKPERMLDRAQLMGLTGPMTVLGGMRVLGTNYGTTKHGVTDCGQLNDPRVNL 674
 Db 600 YMA-PEALVDRAQLGLSGPELTVLVGGURVVGANDSGKDGVTFRNPALSNDFVNL 658
 QY 675 TDMGNSMKPYGVSNAIETDRKRTGAVKWTASRVLDVFGSNLSLRSAEYVAQDDNGEKFEV 734
 Db 659 LSNMTTWSPTAANAFAGHDKRSSEPRWTATRVDLIFGSHAEALRAFAEYACADSQEKFCV 718
 QY 735 DFVAANTKVMNDRFDVAS 753
 Db 719 DFVTAMKVMNADRLDLAA 737

OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2) / M145;
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RA "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 CC 1- FUNCTION: BIFUNCTIONAL, EXHIBITING BOTH A CATALASE AND
 CC BROAD-SPECTRUM PEROXIDASE ACTIVITIES (BY SIMILARITY).
 CC 1- CATALYTIC ACTIVITY: 2 H(2)O(2) -> O(2) + 2 H(2)O.
 CC 1- COFACTOR: Binds 2 protoheme IX and 2 iron ions per tetramer (By
 CC similarity).
 CC 1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL
 CC PEROXIDASE/CATALASE SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AL939106; CAB57412.1; .
 DR HSSP; P00431.1A2F.
 DR InterPro; IPR000763; Bac_ctase/prase.
 DR InterPro; IPR02016; Peroxidase.
 DR Pfam; PF00141; peroxidase; 1.
 DR PRINTS; PR00458; PEROXIDASE.
 DR TIGRFAMs; TIGR00198; cat_per_HPT; 1.
 DR PROSITE; PS00435; PEROXIDASE_1; 1.
 DR PROSITE; PS00436; PEROXIDASE_2; 1.
 DR PROSITE; PS0873; PEROXIDASE_4; 1.
 DR Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide;
 KW Complete proteome.
 FT ACT_SITE 105 BY SIMILARITY.
 FT ACT_SITE 109 BY SIMILARITY.
 FT METAL 272 IRON (HEME AXIAL LIGAND).
 SQ SEQUENCE 740 AA; 80815 MW; 767C0A9BC9BD7798 CRC64;
 Query Match 53.0%; Score 2149.5; DB 1; Length 740;
 Best Local Similarity 55.5%; Pred. No. 6.2e-143;
 Matches 416; Conservative 111; Mismatches 192; Indels 31; Gaps 12;
 24 VSPRHAIIRERAMSGK---CPVMHG--GNTSTGTSNKDWPELNLDILHQDQKSDPMD 78
 1 MSENHDAIVTDARTEEDTDCGVAHGRAPHTQGGNQRQWPERLNKILAKNPANVPLG 60
 79 PDNYIEEVKLDLFDALKDVIHMTDSEWPPADGHGGLMIRNAHWSAGTYRTADGR 138
 61 EEFDYAEAFALDLAAVKRDIAEVLTSQDMWPAFGNGYGLMIRNAHWSAGTYRTISDGR 120
 139 GGGTGSQRFAPLNSPDVSLDKARRLLWPDKKKYGNKISWADLMILAGTVAYESMGLP 198
 121 GGAGAGQQRFAPLNSPDNGNDLKKARRLLWPVAKKYQGNLSWADLVLTGNVALETWGE 180
 199 ATGFSFGYVDIWEPEKDIYWGDEKEWLAPSDERY-GDVNKPETMENPAAVQMGIIYVNP 257
 181 TFCFAGGRADVWEAEEDVYWGPTTWL--DDRYTGD---RELENPLGAVQMGIIYVNP 234

QY 258 EGVNGHPDPLRTAQVLETFARMAMNDEKTAALTAGGHTVGNCHGNASALAPDPKASD 317
 DB 235 EGPNGNDPIAAARDIRETFRMAMNDEETVALIAGGHTFGKTHAGPAGDADVGDDPEAAA 294
 QY 318 VENOGLGWGNPNMOGKASNAVTSIGTAWTTPTKFDMGYEDLLFCYNELKSPAGAHH 377
 DB 295 MEQQLGWKSTHGTGKGDAIYISGLVEVWTSTPTQWNGFFNLFEEFELQSPAGANQ 354
 QY 378 WEPIDIKENKPV--DASDPSIRHNPIMTDADMAIKVNPYRAICFKFADPEYFKKTF 435
 DB 355 W-----VAKDAPEIIPDAHPAKKHPRLMTLTDLSRLDPLYGPISRFFYENPEEPADAF 410
 QY 436 KAWFKLTHDLGPKSRYGPEPAEDLIWQDPI--PAGNTDYCEEV--VKQIAQSGLSI 491
 DB 411 RAWEKLLTHRDMGPKSLYLGPVEETLIWQDPLPEPEGEVIDAEDVATLTKLLESGLSV 470
 QY 492 SEMVSTANDSARTYGRGDMRGANGARIRLAPONEGNEPRERLAKVLVSYE----OISA 547
 DB 471 SOLVITAWASASTFRGSDRGANGARIRLEQRGWENVEPDELAQVLRVLEGVQREFNS 530
 QY 548 DTGA---STADVIVLAGSVGIEKAAKAGYVVRVPELKGKRGDATAEMTDADSFAPLEPLA 604
 DB 531 GSGAKKVSADLIVLGSSAAVEKAAKEAGFPVEVPPAAGRVDATEBHTDAESPEALEPTA 590
 QY 605 DGFERNQKKEYVVKPEMLLDRAQLMGLTGPENTVLLGGMVLTGTYGKTKHGVTDCG 664
 DB 591 DGFERNYLGKGNRLPAEFLLDLRANLLTSLAPEKTVLVGLRVLGAGHQOQSOLGVFTTFC 650
 QY 665 QLTNDFEVLNLTDMGNSKPVGSN--AYEIRDRKTGAVKWTASRVDLVFGSNLLRSYAEV 722
 DB 651 SUTNDFEVLNLTDLGTTWKSTSEDRITTFEGRDAATGEVKWAGSRADLVFGSNLRLAEV 710
 QY 723 YAODDNGEKFRDFFVAAMTKVMNADRFDA 752
 DB 711 YASDDAGEKFXVHDFVAAMVKVNNLDREFLA 740
 RESULT 6
 CATA_ECOLI STANDARD; PRT; 726 AA.
 AC P13029;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Peroxidase/catalase HPI (EC 1.11.1.6) (Catalase-peroxidase)
 DE (Hydroperoxidase I).
 GN KATG OR B3942.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=88314956; PubMed=3045098;
 RA Triggs-Raine B.L., Doble B.W., Mulvey M.R., Sorby P.A., Loewen P.C.;
 RT "Nucleotide sequence of katG, encoding catalase HPI of Escherichia
 coli.";
 RL J. Bacteriol. 170:4415-4419(1988).
 RN [2]
 RP SEQUENCE OF 1-339 FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=93347969; PubMed=8346018;
 RA Plunkett G. III, Burland V.D., Daniels D.L., Blattner F.R.;
 RT "Analysis of the Escherichia coli genome. III. DNA sequence of the
 region from 87.2 to 89.2 minutes.";
 RL Nucleic Acids Res. 21:3391-3398(1993).
 RN [3]
 RP SEQUENCE OF 309-736 FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=94089392; PubMed=8265357;
 RA Blattner F.R., Burland V.D., Plunkett G. III, Sofia H.J.,
 RA Daniels D.L.;
 RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the

QY	360	LLFGYNWELKKSPAGAHHPEDIDIKENKFPVDASDPSIRHNPIMTDADMAIKVNPYRAI	419
Db	332	NLFKYMVQTRSPAGAIQEAVD-APEIIP-DPFDSEKKRPTMLVTDLTLPDPPEFKI	389
QY	420	CEKFMADPPEYFKKTKPAKWFKTHRDGLPKRSYIGPEVPAEDLIWODPIPAQNTDYCEEV	479
Db	390	SRRELTNDPOAFNEAFARAWFKLTHRDGPKRSYIGPEVKEDLIWQDPLPQPIYNPTEQD	449
QY	480	V--KOKIAQSGLSISEMYSTANDSARTYRSGSDMRGANGARILRAPQNEWOGNEPERLA	536
Db	450	ITDLFAIADSGLSVSELSVAVASASTFRGGDKGGANGARLALMPQRDWDVN--AAAV	507
QY	537	KVLSYVEQISADTG-ASIAIDVTVLAGSVGIEKAAAGYDVVPFLKGRGDATAEWTDAD	595
Db	508	RALPVLKRIQKESGRASLADIIVLAGVGVGEKASNAAGLSIHVPFAPGRVARDARQODTIE	567
QY	596	STAPLEPLADGFRNMOKKEVYVVKPEEMLLDRAQLMGLTGPEMTVLGGMRVLGTNYGSK	655
Db	568	MFELLEPIADGFRNTRARLDVSTTESLLDKAQQLTLTAPENTALVGGMRVLGANFDGSK	627
QY	656	HCVFTDCSGOLTNDFVNLTDMGNSKKPV--GSNAYEIRDRKTGAVKWTASRVDLVFGSN	713
Db	628	NGVFTDRVGLSNDFFVNLLDMRYEMKATDESKELFEGDRDETGEYKFTASRADLVFGSN	687
QY	714	SLLRSGVAYEAOODNGEKFVRDFVAAWTKVMNADRFDV	751
Db	688	SVLRVAEYAYSSDAHEKFEKDFVAAWTKVMNLDREFDL	725
RESULT 7			
CATA_SALTY			
ID	CATA_SALTY	STANDARD;	PRT; 726 AA.
AC	P17750;		
CD	01-AUG-1990 (Rel. 15, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	15-SEP-2003 (Rel. 42, Last annotation update)		
DE	peroxidase/catalase HPI (EC 1.11.1.6) (Catalase-peroxidase)		
DE	(Hydroperoxidase I).		
GN	KATG OR STM4106.		
OS	Salmonella typhimurium.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Salmonella.		
OX	NCBI_TaxID=602;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=LT2;		
RC	MEDLINE=911117169; PubMed=2277629;		
RA	Loewen P.C., Stauffer G.V.;		
RT	"Nucleotide sequence of katG of Salmonella typhimurium LT2 and		
RT	characterization of its product, hydroperoxidase I.;"		
RL	Mol. Gen. Genet. 224:147-151(1990).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=LT2 / SGSC1412 / ATCC 700720;		
RC	MEDLINE=21534948; PubMed=11677609;		
RA	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,		
RA	Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,		
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,		
RA	Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,		
RA	Waterston R., Wilson R.K.;		
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium		
RT	LT2.;"		
RL	Nature 413:852-856(2001).		
CC	-!- FUNCTION: Bifunctional, exhibiting both a catalase and		
CC	broad-spectrum peroxidase activities.		
CC	-!- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.		
CC	-!- COFACTOR: Binds 2 prothème IX and 2 iron ions per tetramer.		
CC	-!- SUBUNIT: Homotetramer.		
CC	-!- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL		
CC	PEROXIDASE/CATALASE SUBFAMILY.		
CC	-!- SIMILARITY: Contains 1 heme regulatory motif (HRM) repeat.		

This SWISS-PROT entry is copyright. It is produced through a collaboration

This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AL627279; CAD09515.1; -
DR EMBL; AE016846; AA071018.1; -
DR InterPro; IPR000763; Bac_ctase/prase.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00141; peroxidase; 1.
DR PRINTS; PR00458; PEROXIDASE.
DR TIGRFAMS; TIGR00198; cat_per_HPI; 1.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS00436; PEROXIDASE_2; 1.
DR KX Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide;
KW Complete proteome.
FT ACT_SITE 102 102 BY SIMILARITY.
FT ACT_SITE 106 106 BY SIMILARITY.
FT METAL 267 267 IRON (HEME AXIAL LIGAND).
FT REPEAT 15 20 HRM.
SQ SEQUENCE 726 AA; 79597 MW; B701C95343EEBB38 CRC64;

Query Match 51.3%; Score 2083; DB 1; Length 726;
Best Local Similarity 53.6%; Pred. No. 2.8e-138;
Matches 406; Conservative 102; Mismatches 206; Indels 44; Gaps 9;

QY 4 ASADLHSSLQORCRAFVPLVSPRHRAIRERAMSGKCPVMHGG---NTSTGTSNKDWPE 60
DB 2 STTDDTHNTLS-----TGKCPFFQGGHDSAGAGTAGRDWPN 39
QY 61 GLNLDILHQQRKSDPMDPPFNRYEYVKLDFALKDVHALMTDSQEMPADGHYGG 120
DB 40 QLRVDLLNQHSNRNPLGDFDYKESKLDYSALKGLKALLTDSQPNWADWGSYVL 99
QY 121 MIRMAHSACTYRTADRGCGGTCGSRFAPLNSPDVSLDKARRLLWPIKKYKNSIW 180
DB 100 FIRMAHGAITYRIDRGAGRGQQRAPLNSPDVSLDKARRLLWPIKKYKNSIW 159
QY 181 ADLMILAGTVAYESMGLPAYGFSRVDIWEPEKDIYWGDEKEWKLAPSDERYGVNKPET 240
DB 160 ADLFILAGNVALENSGRTFGAGREDVWEPDLVDVNGDEKALWTH-----RIPEA 211
QY 241 M-ENPLAAVQMLIYVNEGVNHPDLRTAQOVLFTFARMNDEKTAALTAGHHTVGN 299
DB 212 LAKAPLGATEMLIYVNEGVNHPDLRTAQOVLFTFARMNDEKTAALTAGHHTVGN 271
QY 300 CHGNGNASALAPDKASDVENOGIWCNPNWQKASNAVTSIGEGANTTPTKFDGMYFD 359
DB 272 THGAAAASHGVGADEAAPIEAQGLGWASSYGSYGVDADITSGLEVWVTQTOWSNYFFE 331
QY 360 LLFGYNWELKKSAPAGHHEWPEIDIKKKNKVPDASDPSIRINPIMTDADMAIKVNPYRAI 419
DB 332 NLKFEYVQVRSAGALQFAVD---APDIIPDFDPSKPKPMLVTLTDLRFDPEPEKI 389
QY 420 CEKFMADPEYKTKFAKAWFKLTHRDIGPKSRVIGPEVPAEDLIWQDDIPAGNTDYCEE 479
DB 390 SRRLFLNDPOAFNEAFARAWFKLTHRDIGPKSRVIGPEVPAEDLIWQDDIPAGNTDYCEE 449
QY 480 V---KQIAQSGLSISEMVSTANDSARTYRGSDMRGCGANGARILAPONTWQGNPERLA 536
DB 450 IINKAAIAASGSISEMSVWASASTFRGGKRGANGARILAPONTWQGNPERLA 507
QY 537 KVLSTVEQISADTG-ASIAIVIVLAGSVGTEKAAAGYDVRVPLFKGRGDATMTDAD 595
DB 508 RVLPLVLEALQKTNKASLADIIVLAGVVGLEQAAAAGVSIYFPAPGRVADARQDDIE 567
QY 596 SFAPLEPLADGFRNQKKEYVPEEMLLDRAQLMGLTGPEVLLGGMRVLTGNYGKTK 655
DB 568 MFSLLLEPLADGFRNRYARLDVSTTESLLIDKAAQQLTAPETMIVLVGMRVLTGNYGKTK 627
QY 656 HGVTDCGQLTNDFFVNLTDGMSNKPV---GSNAYEIRDKTKGAVKWTASRDVLVFGSN 713

DB 628 NGVETDRPGVLTSTOFFANLLDMRYEWKPTDDANELFEGDRDLTGCVKYATRADLVFGSN 687
QY 714 SLRSYAEVYAADONGKEKVFDFVAANTVKVNAADRFV 751
DB 688 SVLRALAEVYACSDAHEKEKVFDFVAANTVKVNAADRFV 725

RESULT 9
ID CATA_MYCIT STANDARD; PRT; 746 AA.
AC Q04657;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Peroxidase/catalase (EC 1.11.1.6) (Catalase-peroxidase) (M185 protein).
DE protein.
GN M185.
OS Mycobacterium intracellulare.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID:1767;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-93123988; PubMed-1336034;
RA Morris S.L., Nair J., Rouse D.A.;
RT "The catalase-peroxidase of Mycobacterium intracellulare: nucleotide sequence analysis and expression in Escherichia coli.";
RL J. Gen. Microbiol. 138:2363-2370(1992).
CC -1- FUNCTION: BIFUNCTIONAL, EXHIBITING BOTH A CATALASE AND BROAD-SPECTRUM PEROXIDASE ACTIVITIES. MAY PLAY A ROLE IN THE INTRACELLULAR SURVIVAL OF MYCOBACTERIA.
CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) -> O(2) + 2 H(2)O.
CC -1- COFACTOR: Binds 2 prothème IX and 2 iron ions per tetramer.
CC -1- SUBUNIT: Homotetramer (Probable).
CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL PEROXIDASE/CATALASE SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----

DR EMBL; M86741; AAA25360.1; -
DR HSP; P00431; 2CYP.
DR InterPro; IPR000763; Bac_ctase/prase.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00141; peroxidase; 1.
DR PRINTS; PR00458; PEROXIDASE.
DR TIGRFAMS; TIGR00198; cat_per_HPI; 1.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS00436; PEROXIDASE_2; 1.
DR PROSITE; PS00873; PEROXIDASE_4; 1.
KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.
FT ACT_SITE 110 110 BY SIMILARITY.
FT ACT_SITE 114 114 BY SIMILARITY.
FT METAL 277 277 IRON (HEME AXIAL LIGAND).
SQ SEQUENCE 746 AA; 81417 MW; 15F35F7F5028F2B2 CRC64;

Query Match 50.9%; Score 2063.5; DB 1; Length 746;
Best Local Similarity 55.2%; Pred. No. 6.7e-137;
Matches 395; Conservative 98; Mismatches 197; Indels 25; Gaps 10;

QY 53 SNKDWPEGNLILHQQRKSDPMDPPFNRYEYVKLDFALKDVHALMTDSQEWPA 112
DB 40 TNRDWPQDVQVSSLPHSPNSPLNGDFFDYAAAEKALDVEALKADMISLMTTSQDWPA 99
QY 113 DWGHYGLMIRMAHWSAGTYRIADRGCGGTCGSRFAPLNSPDVSLDKARRLLWPIKK 172
DB 100 DYGHYGLFIRMSHAAGTYRIADRGCGGTCGSRFAPLNSPDVSLDKARRLLWPIKK 159

173 KYGNKISWADLMILAGTVAESMGLPAYGFSRVDIWEPEKDIWNGDEKEWLAPSDERY 232
 160 KYGNKISWADLIYAGNALESMSGKTFGFGREDVWEPE-EILWGEDEEWLG-TDKRY 217
 233 GDVNPETWENPLAIVNPGVNGVHDDPLRTAOOVLETARMANDEKTAALTA 292
 218 SGERE---LAQPYGATTGLIIVNPEGEGKDDPATAADIRRTFGMANDEEATAALIV 274
 293 GGHVTGNGCHNGNASALADPKKASDVENQGLGNGNPMQGNASNAVTSIGEGAWTNPTK 352
 275 GGHSGFKTHGAGDADLVGEPEAAPIEQOGLGKSSYGTSGKDAITSLGVVWTPPK 334
 353 FDMGTFDLLFGYNWELKSPAGAHWEPIIDIKENKPYDASP--SIRINPIMTDADMAI 410
 335 WDNSELETLYGVEWELTKSPAGA--WQ-PTAKDGAGAGTIPDPFGGAGRAPTMLVTDISL 391
 411 KYNPTYRAICEKFMADPEVFKTFKAMFKLTHRDLPKRSYIGPEVPAEDLIWQDPIPA 470
 392 RESPIADITRRWLDHPEELADAFKAMTKLHRDMGPISRYLGPWV-AEPLWQDPPVA 450
 471 GNTDYCEE---VVKOKIAQSGLSISEMSTAWDSARTYRGSMDRGGANGARIRLAPQNE 526
 451 VDBELVDDNDVAALKKVLDSGLSIPQLVKTASAAASYRNTDKRGANGRLRLOPQRS 510
 527 WOGNEPERLAKVLSYVEQISADTGA-----SIADVIVLAGSVGIEKAAKAAGYDVRV 578
 511 WEVNEPSELDKALPVLEKIQODFNASAGGKKISLADLIVLAGSAVEKAAKADAGYEISV 570
 579 PFLKRGDGTAEWTDADSPAPLEPLADGPRNQKKEYYVVKPEMLIDRAQLMGLTGPENT 638
 571 HFAPGRTDASQESTDVSFAVLEPRADGFRNIRPGEKAPLEQLLIERAYLILGVTPGEMT 630
 639 VLLGGMRLVGTNYGGTKHGVTDCGQLTNDFFVNLDMGNSWK--PVGSNAYEIRDRKT 696
 631 VLVGGLRALGNHSGKHGVFTDRPGALTNDFVNLDMGTENKASETAENYVEGRDRAS 690
 697 GAVKWTASRVLDVFGSNLLRSYAEVYAADNGEKFVRDFAVWTKVNNADREDV 751
 691 GALKWTATANDLVFGSNVLRGLVEYAADHAKFVEDFVAWVKVMSNDRFDL 745

RESULT 10
 CATL_MYCFO STANDARD; PRT: 752 AA.
 AC O08404;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Peroxidase/catalase 1 (EC 1.11.1.6) (Catalase-peroxidase 1).
 GN KATGL.
 OS Mycobacterium fortuitum.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1766;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 6841;
 RX MEDLINE=98037481; PubMed=9371430;
 RA Menendez M.C., Alnsa J.A., Martin C., Garcia M.J.;
 RT "katI and katII encode two different catalases-peroxidases in
 RT Mycobacterium fortuitum.";
 RL J. Bacteriol. 179:6880-6886(1997).
 CC -I- FUNCTION: BIFUNCTIONAL, EXHIBITING BOTH A CATALASE AND
 CC BROAD-SPECTRUM PEROXIDASE ACTIVITIES. MAY PLAY A ROLE IN THE
 CC INTRACELLULAR SURVIVAL OF MYCOBACTERIA (BY SIMILARITY).
 CC -I- CATALYTIC ACTIVITY: 2 H₂O(2) - O(2) + 2 H₂O.
 CC -I- COFACTOR: Binds 2 prothème IX and 2 iron ions per tetramer.
 CC -I- SUBUNIT: Homodimer (Probable).
 CC -I- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL
 CC PEROXIDASE/CATALASE SUBFAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; Y07865; CAA69192.1; -
 DR HSSP; P00431; ICCK.
 DR InterPro; IPR000763; Bac_ctase/prase.
 DR InterPro; IPR002016; Peroxidase.
 DR Pfam; PF00141; peroxidase; 1.
 DR PRINTS; PR00458; PEROXIDASE.
 DR TIGRFAMs; TIGR00198; cat_per_HPI; 1.
 DR PROSITE; PS00435; PEROXIDASE_1; 1.
 DR PROSITE; PS00436; PEROXIDASE_2; 1.
 DR PROSITE; PS00873; PEROXIDASE_4; 1.
 KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.
 FT ACT_SITE 113 BY SIMILARITY.
 FT ACT_SITE 117 BY SIMILARITY.
 FT METAL 285 IRON (HEME AXIAL LIGAND).
 SQ SEQUENCE 752 AA; 83100 MW; B66BAD53DD26E70D CRC64;

 Query Match 50.5%; Score 2050; DB 1; Length 752;
 Best Local Similarity 54.3%; Pred. No. 5.9e-136;
 Matches 388; Conservative 102; Mismatches 203; Indels 22; Gaps 8;
 QY 53 SNKDWPEGLNLDILHQDKRSDPMDPFDNYREVRKLDLFDALKKDVHMLTDSQWPPA 112
 DB 43 TWQDWPDQDVSVRLHKQIEGNPLGAGFNAYAEFOKLDVEALRADMLELMTSSQWWR 102
 QY 113 DMGHVGLMIRMAWHSAGTYRIADRGGGGTGSGRFAPLNSPDNYSLKARLLWPICK 172
 DB 103 DYGTVAGLSTRSWHAAGTYRIPDGRGGAGQAGREAPINSPDNYSLKARLLWPICK 162
 QY 173 KYGNKISWADLMILAGTVAESMGLPAYGFSRVDIWEPEKDIWNGDEKEWLAPSDERY 232
 DB 163 KYGNKISWADLIIFAGNVALESAGTFTGFGAQDIWEPE-EILWGEDEEWLG-TDKRY 220
 QY 233 GDVNPETWENPLAIVNPGVNGVHDDPLRTAOOVLETARMANDEKTAAL 290
 DB 221 GGTNDSTNRELANPYCATTMGLIIVNPEGEGKDDPATAADIRRTFGMANDEEATAAL 280
 QY 291 TAGGHTVGNCHNGNASALADPKKASDVENQGLGNGNPMQGNASNAVTSIGEGAWTNP 350
 DB 281 IVGGHTLGTHGPGDLVGEPEAAPIEQOGLGKSCAFSGRGSDDTITSGLEVWVWTPP 340
 QY 351 TKFDMCYFDLLFGYNWELKSPAGAHWEPIIDIKENKPYDASDPSIRHNPIMTDADMAI 410
 DB 341 TWNSNYLEILYGEWELTKSPGDAWQFKAADAEIIPDPFGGPP---RKPTMLVTDISM 397
 QY 411 KYNPTYRAICEKFMADPEVFKTFKAMFKLTHRDLPKRSYIGPEVPAEDLIWQDPIPA 470
 DB 398 RVDPIYGPITRRWLEHPEELNEAFKAMTKLHRDMGPISRYLGPWIP-EPQLWQDPPVD 456
 QY 471 GNTDYCEE---VVKOKIAQSGLSISEMSTAWDSARTYRGSMDRGGANGARIRLAPQNE 526
 DB 457 VDHPLVDEQDIAALKKLLDLSGLSVQQLVKTASAAASFRGTDKRGANGRLRLOPQRN 516
 QY 527 WOGNEPERLAKVLSYVEQISADTGA-----SIADVIVLAGSVGIEKAAKAAGYDVRV 578
 DB 517 WEVNEPSELDKALPVLERIAAQDFNASASDGKKISLADLIVLAGSAIEKAAADGGYEVK 576
 QY 579 PFLKRGDGTAEWTDADSPAPLEPLADGPRNQKKEYYVVKPEMLIDRAQLMGLTGPENT 638
 DB 577 HFVAGTDSQENTDVSFAVLEPRADGFRNIRPGEKAPLEQLLVDKATFLNLTAPMT 636
 QY 639 VLLGGMRLVGTNYGGTKHGVTDCGQLTNDFFVNLDMGNSWK--VGSNAYEIRDRKT 696
 DB 637 VLVGGLRALMTHGSGKHGVFTANPGALSNDFFVNLDMSTENKPSETAENYVEGRDRRT 696
 QY 697 GAVKWTASRVLDVFGSNLLRSYAEVYAADNGEKFVRDFAVWTKVNNADREDV 751
 DB 691 GALKWTATANDLVFGSNVLRGLVEYAADHAKFVEDFVAWVKVMSNDRFDL 751

Db 298 LQWSSYGTGKDAITSGIEVWVWNTPTKWDNSFLEILLYGYEWELTKSPAGAWQYAKD 357
 QY 303 -IKKENKPVADSDPISIRINPIMTDADMAIKVNPYTRAIKCEKPMADPEYFKTKFAKAWFKL 441
 Db 358 GAGACTIPDPFGPG--RSPTMLATDLSRLVDPIYERITRRWLEHPEELADEFAKAWYKL 415
 QY 442 THRDIGPKSRIGYRVPVPAEDLIWQDPIAGNTDYCEEV-----VKOKIAQSGLSISEMYST 497
 Db 416 IHRDMGPVARYLGLPVLVQKTLWQDPVPAVSHDLVGEAEIASLKSQIILASGLTVSOLYST 475
 QY 498 AWDSARTYRGSDMRGANGARIRLAPQNEWOGNEPE--BLAKVLSYEQI-----SADYK- 550
 Db 476 AWAASSPFGSDKRGANGRIRLQPOQVGEWVNDPDGLRKVIRTEIIEQSFNSAAPGN 535
 QY 551 --ASTADVIVLAGSVIGIEKAKAGYDVRVFLPKRGDATAEMTDADSFAPLEPADGFR 608
 Db 536 IKVSFADLVVLGGCAIEKAKAAGHNITVPTPGRTDASQEQTDVESFAVLEPKADGFR 595
 QY 609 NWOKKEYVVKPEEMLLDRAQLMGLTGPEMTVLGGMRVLTGNYGCTKHGVETDCGQLTN 568
 Db 596 NYLKGPNLPAYMLLDKANLTLSAPEMTVLVGLRVLGANYKRLPLGVFTEASESLTN 655
 QY 669 DFFVNLTDMGNSK--PVGSNAYEIRDRKTGAVKWTASRVDLVFGSNLSLSYAEVYAAD 726
 Db 656 DFFVNLLDMGITWEPSPADDTGYQKD--GSGKVKWTGSRVDLVFGSNLSRLALVEYQAD 714
 QY 727 DNGEKFVDFVAATKVMNADRFV 751
 Db 715 DAQPKFVQDFVAAMDVMNLDREFV 739

RESULT 13

CAT2_NEUCR
 ID CAT2_NEUCR STANDARD: PRT: 753 AA.
 AC Q8X182;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Peroxidase/catalase 2 (EC 1.11.1.6) (Catalase-peroxidase).
 GN CAT-2.
 OS Neurospora crassa.
 CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 CC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 CC NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=22029357; PubMed=12033445;
 RA Peraza L., Hansberg W.;
 RT "Neurospora crassa catalases, singlet oxygen and cell
 RT differentiation.";
 RL Biol. Chem. 383:569-575(2002).
 CC -1- FUNCTION: Bifunctional, exhibiting both a catalase and
 CC broad-spectrum peroxidase activities.
 CC -1- CATALYTIC ACTIVITY: 2 H₂O₂ - O₂ + 2 H₂O.
 CC -1- COFACTOR: Protoheme IX and iron (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL
 CC PEROXIDASE/CATALASE SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF459787; AAL66352.2; -
 CC InterPro: IPR002016; Peroxidase.
 CC Pfam; PF00141; peroxidase; 1.
 CC PRINTS; PR00458; PEROXIDASE.
 CC PROSITE; PS00435; PEROXIDASE_1; 1.
 CC PROSITE; PS00436; PEROXIDASE_2; 1.
 CC PROSITE; PS50873; PEROXIDASE_4; 1.

KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.
 FT ACT_SITE 87 BY SIMILARITY.
 FT ACT_SITE 91 BY SIMILARITY.
 FT METAL 279 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
 SQ SEQUENCE 753 AA; 83379 MW; DE01DC6D10BA582 CRC64;

Query Match 49.4%; Score 2003; DB 1; Length 753;
 Best Local Similarity 50.9%; Pred. No. 1.2e-132;
 Matches 382; Conservative 121; Mismatches 199; Indels 48; Gaps 10;

QY 39 KCPVHGSGNTSGTSNKKDWPEGLNLLHQQDRKSDPDMDPENTYREEVRKLDLFDALKKD 98
 Db 3 ECPVKSNSVGGGTRHNDWMPAQLRLNLRQPTVPSNPLDKDFDFAAFKSLDGLKKD 62
 QY 99 VHALMTDSQEWMPADWGHYGGMLIRMAWHSAGTYRIADCGRGGGTSGSORFAPLNSWPDNV 158
 Db 63 LTKLMTDSQDWMPADFGHYGGLFIRMAWHSAGTYRVTDRGGGGGQGRFAPLNSWPDNV 122
 QY 159 SLDKARRLLWPKKKYGNKISWADLMILAGTVAYESMGLPAYGFSGRVDIWEPEKDIY 218
 Db 123 SLDKARRLLWPKKKYGNKISWADLMILAGTVAYESMGLPAYGFSGRVDIWEPEKDIY 218
 QY 219 GDEKEMLAPSDERY-----GDVKNKE-----TMENPLAAVQMGILYVNP 257
 Db 183 GAETTWLG--NDRYSEGGEGHGVGGDESKKQHTDIHNRDLQSLASSHMGILYVNP 241
 QY 258 EGVNGHPDLRTAQOVLTFARMAMNDEKTAALTAGHTVGNCHGNASALAPDPKASD 317
 Db 242 EGPDGIPDVASAKDIRVTFGRMANDEETVALIAGHSFGKTHGAGTHHVGEKEPAAP 301
 QY 318 VENQGLGNGNPMQOKASNAVTSGETAGWTNPTKFDGMYFDLLFCYNWEIKKSPAGAH 377
 Db 302 IEHQGLGWANSFGQKGPDTITSGLEVTWTPPTKMGMYLEYLYKFDWEPTKSPAGANQ 361
 QY 378 WEPIDIKKENPV--DASDPSTRHNPIMTDADMAKVNPTYRAICEKPMADPEYFKTKFA 435
 Db 362 W-----VAKNAEPTIPDAYDPNKKLPTMTDIALRMDPAYDKICRDLANPDKPADAPA 417
 QY 436 KAWFKLTHRDIGPKSRIGYRVPVPAEDLIWQDPIAGNTDYCEE-----VVKOKIAQSGLSI 491
 Db 418 RAWFKLLHRDMGPRTRWIGPEVPSILPWEDYIPVDYQIIDDNDIALKKEILATGAP 477
 QY 492 SEMYSTANDSARTYRGSDMRGANGARIRLAPQNEWOGNEPERLAKVLSYEQI----- 545
 Db 478 KKLIFVWSSASSPFGSDKRGANGARIRLAPQNEWKNVDPTSLREVLAALESVQOKFND 537
 QY 546 -SADTGASIAADVIVLAGSVIGIEKAKAGYDVRVFLPKRGDATAEMTDADSFAPLEPLA 604
 Db 538 SSSGKKVSLADLIIVLGGVALEQAS-----GLVVPFTPCRNDATQEHDTVHSHFLEPHA 592
 QY 605 DGFNRWQKKEYVVKPEEMLLDRAQLMGLTGPEMTVLGGMRVLTGNTYGGTKHGVETDCG 664
 Db 593 DGRSYGKGTGRVTRTEQFLIDRASLLTSLAPELATALIGLRLVLEANYDSSGYVLTTPG 652
 QY 665 QLTNDFVNLDMGNSKWPVGSNAYEI---RDRKTGAVKWTASRVDLVFGSNLSLSYAE 721
 Db 653 KLTNDYFVNLLDTNWK--AADNEGEVFIGYDRKTHDKKTATRADLTADLFGAAHLEALAE 711
 QY 722 VYAADNGEKEFVRDFVAATKVMNADRFV 751
 Db 712 VYAAVDGEKEFVRDFVAATKVMNADRFV 741

RESULT 14

CATA_LEGPN
 ID CATA_LEGPN STANDARD: PRT: 749 AA.
 AC Q9WAB9;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Peroxidase/catalase (EC 1.11.1.6) (Catalase-peroxidase).
 GN KATG.
 OS Legionella pneumophila.

[illegible]

Qy	473	TDY-----CEEV--VKQIAQSGUSISEMYSWSTANDSARTYKSGDMRGANGARIRLAPQNE	526
Dd	446	VDYKLVLDANDIANLKGKILNSGLTTSSELVKTANASASTFRGTDMRGANGARIRLAPQKD	505
Qy	527	WQGNPERLAKVLSVYEQI-----SAD--TGASIADVIVLAGSVGIEKAAKAGYDVRV	578
Dd	506	WPANDPOELAKVILKTLESIONNNNAQADGKISLADLIVLGGNAATEQAAKQAGYDIIV	565
Qy	579	PFILKGRGDATAEMTDDADSFAPLADGFRNWKQKEYVVKVPEMLLDRAQLMGLTGPEMT	638
Dd	566	PFPYGRDTAQTGMDYKSVFLEPKADGFRNFDKSNMSPPEMLVKEKASLLKLSVPEMT	625
Qy	639	VLLGGMVILGTNGTGGTKHGVFTDCGGLTNDNFVNTLDMGNSKWPVG--SNAYEIRDRKT	696
Dd	626	VLVGGMRVLNANTGQNGYGVFTDKPGPLNDDFFINLLSMSTENKKSSETGIEGYEVRKT	685
Qy	697	GAYKWTASRDVLVFGSNLLRSYAETVAODDNGEKFVRDFFAAWTKVMNAADRFDV	751
Dd	686	GKLWKATSYDLIFGANSELRVAEAYATODAKEFKIQDFINAWKVTADREFDI	740
RESULT 15			
ID	CATA_HALMA	STANDARD;	PRET; 730 AA.
AC	OS9651;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	15-SEP-2003 (Rel. 42, Last annotation update)		
DE	Peroxidase/catalase (EC 1.11.1.6) (Catalase-peroxidase).		
GN	PERA.		
OS	Haloarcula marismortui (Halobacterium marismortui).		
OC	Archaea; Euryarchaeota; Halobacteriia; Halobacteriales;		
OC	Halobacteriaceae; Haloarcula.		
OX	NCBI_TaxID=2238;		
EN	[1]		
RP	SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND MASS SPECTROMETRY.		
RX	MEDLINE=9912366; PubMed=9924978;		
RA	Cannac-Caffrey V., Hudry-Clergeon G., Petillot Y., Gagnon J.,		
RA	Zaccari G., Franzetti B.;		
RT	"The protein sequence of an archaeal catalase-peroxidase.";		
RL	Biochimie 80:1003-1011(1998).		
CC	-1- FUNCTION: BIFUNCTIONAL, EXHIBITING BOTH A CATALASE AND		
CC	BROAD-SPECTRUM PEROXIDASE ACTIVITIES (BY SIMILARITY).		
CC	-1- CATALYTIC ACTIVITY: 2 H(2)O(2) -> O(2) + 2 H(2)O.		
CC	-1- COFACTOR: Binds 2 prothème IX and 2 iron ions per tetramer (By		
CC	similarity).		
CC	-1- MASS SPECTROMETRY: MW=81292; MW-ERR=9; METHOD=MALDI		
CC	-1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL		
CC	PEROXIDASE/CATALASE SUBFAMILY.		

CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		

DR	EMBL: Y16851; CAA76423.1; -		
DR	PIR: T44846; T44846.		
DR	PDB: 1ITK; 28-AUG-02.		
DR	InterPro: IPR000763; Bac_ctase/prase.		
DR	InterPro: IPR002016; Peroxidase.		
DR	Pfam: PF00141; peroxidase; 1.		
DR	PRINTS: PR00458; PEROXIDASE.		
DR	TIGRFAMS: TIGR00198; cat_per_HPI; 1.		
DR	PROSITE: PS00435; PEROXIDASE_1; 1.		
DR	PROSITE: PS00436; PEROXIDASE_2; 1.		
DR	PROSITE: PS50873; PEROXIDASE_4; 1.		
DR	Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide;		
KW	3D-structure.		
FT	INIT_MET 0		
FT	ACT_SITE 91		BY SIMILARITY.

```
FT ACT_SITE 95 95 BY SIMILARITY.
FT METAL 258 258 IRON (HEME AXIAL LIGAND).
SQ SEQUENCE 730 AA; 81252 MW; 5C7EFE6C37ADC4B CRC64;

Query Match 49.28; Score 1998; DB 1: Length 730;
Best Local Similarity 52.78; Pred. No. 2.5e-132;
Matches 385; Conservative 103; Mismatches 219; Indels 24; Gaps 8;

QY 36 MSGKCPVMHGGNTSTGTSNKDMPEGLNLILHQQDKSDPMDPDFNYREVRKLDLFDAL 95
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8 MSGAT---GGRSKRPSNQDWPSKLNLEILQONARDVGPVEDDFDYAEFOKLDLEAV 63

QY 96 KKDVAHALTDSQWHPADWGHYGLMIRMMAHWSAGTYRIADGRGGGTGSGRQFAPLNSWP 155
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 64 KSDLEELMTSSQDWHPADYCHYGFLFIRMAHWSAGTYRTADGRGAAGGRQREAPINSWP 123

QY 156 DNVSLDKARLLWPIKKYCNKISWADLMTLACTVAVESMGLPAYCFSGRVDIWEPEKD 215
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124 DNANLQKARLLWPIKKYCNKISWADLMTLACTVAVESMGLPAYCFSGRVDIWEPEKD 183

QY 216 IYWGDEKELWAPSDERYGVNKPETMENPLAAVQMGIIYVNBGVNGHPDPLRTAQOYLE 275
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 184 VNGPDEPF--ETQERF---DEPGEIQEGLGASVMGLIYVNBGVNGHPDPLRTAQOYLE 238

QY 276 TFRMAAMNDEKTAALTAGHTVNCNGHGNASA-LAPDPKASDVENOGGLGWNPNMGKA 334
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 239 TFRMAAMNDEKTAALTAGHTVNCNGHGNASA-LAPDPKASDVENOGGLGWNPNMGKA 334

QY 335 SNAVTSIGCAWTNTPKFDMGYFDLLFGYNWELKKSPAGAHWEPIIDIKKENKPVDAKD 394
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 299 GEMITSGIEGFWTQSTEDWDMGINLLDYEWEPKPGPGGAWQWAPKSEELKNSVDAHD 358

QY 395 PSIRHPIMTDADMAIKVNPYTRAIACEKNADPEYFKTFPAKAWFKLTHRDGLGPKSRYIG 454
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 359 PDEKQTPMMLTTDIALKRPDPYREVMTFOENPMFCMNFKAAYKLTHERDMGPPERFLG 418

QY 455 PEYPAEDLIWQDPIPAQNTDYC--BEV--VKQIAOSGLSISEMVSATWDSARTYRGSDM 510
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 419 PEYPAEDLIWQDPIPAQNTDYC--BEV--VKQIAOSGLSISEMVSATWDSARTYRGSDM 510

QY 511 RGGANGARIRLAPQNEWQNEPERLAKVLSVYEQISAD-----TGASIAADVIVLAGS 562
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 479 RGGANGARIRLAPQNEWQNEPERLAKVLSVYEQISAD-----TGASIAADVIVLAGS 562

QY 563 VGIEKAAAGYDVRVPFLKGRGDATAMTDADSFAPLEPLADGFRNMQKKEYVVRPEEM 622
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 539 AAVEQAAANAGYDVEIPEFGRVDAGPEHTDAPFDALKPKVDGVARNYIQDDITRPAEEV 598

QY 623 LLDRAQLMGLTGPMVTLGGMRVLTGNYGGTKHGVTDCGGLTNDFFVNLDMGNSWK 682
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 599 LVDNADLLNLTASELTAIGMRSIGANYQDTDLGVFTDEPETLTNDFFVNLDMGTEWE 658

QY 683 PVGSNAYEIR--DRKTKAVKWTASRYDLVFGSNLLRSYAETVAQDDNGEKFYRDFVAW 740
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 659 PAADSEHRYKGLDRDTGEVKNWEATRIDLIFGSDNRRLRAISEVYGSADAERKLVHDFVDTW 718

QY 741 TKVMNADREFDV 751
   ||| ||| |||
Db 719 SKVMKLDREFDL 729
```

Search completed: October 7, 2003, 19:43:00
Job time : 20.5881 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: October 8, 2003, 06:25:58 ; Search time 8206.54 Seconds
(without alignments)
11276.082 Million cell updates/sec
Title: US-09-884-889-5
Perfect score: 2262
Sequence: 1 atgaataacgcacccgctga.....gtttgcagcgcgcgtcgtaa 2262

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pi:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_un:*
28: em_vi:*
29: em_hcg_hum:*
30: em_hcg_inv:*
31: em_hcg_inv:*
32: em_hcg_inv:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rod:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_atgo_hum:*
40: em_atgo_mus:*
41: em_atgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2262	100.0	2262	6	AR098263	Sequence
2	2262	100.0	2262	6	AR215251	Sequence
3	1149	50.8	14482	1	AE004233	Vibrio ch
4	1098	48.5	7772	1	AE008223	Agrobacte
5	1098	48.5	10299	1	AE009392	Agrobacte
6	1096.4	48.5	2640	1	AB033631	Agrobacte
7	1085	48.0	296100	1	AF005086	Vibrio pa
8	1076	47.6	2700	1	AF197161	Synechoco
9	1073	47.4	304490	1	AF016806	Vibrio vu
10	1031	45.6	303450	1	AP005085	Vibrio pa
11	1030.4	45.6	3865	1	AF078110	Legionell
12	1019.2	45.1	2430	1	SYOGP	D61378 Synechococc
13	951.2	42.1	12293	1	AE015873	Shewanell
14	945.2	41.8	2896	1	SYCKATG	D83990 Synechocyst
15	945.2	41.8	146174	1	D90910	Synechocyst
16	837.2	37.0	2208	1	AB020234	Bacillus
17	837.2	37.0	2232	1	AB020065	Bacillus
18	837.2	37.0	2232	1	AB020120	Bacillus
19	837.2	37.0	2235	1	AB020067	Bacillus
20	837.2	37.0	2235	1	AB020092	Bacillus
21	837.2	37.0	2238	1	AB020115	Bacillus
22	837.2	37.0	2241	1	AB020094	Bacillus
23	837.2	37.0	2247	1	AB020078	Bacillus
24	837.2	37.0	2250	1	AB020106	Bacillus
25	837.2	37.0	2253	1	AB020116	Bacillus
26	837.2	37.0	2253	1	AB020117	Bacillus
27	837.2	37.0	2256	1	AB020064	Bacillus
28	837.2	37.0	2256	1	AB020066	Bacillus
29	837.2	37.0	2256	1	AB020071	Bacillus
30	837.2	37.0	2256	1	AB020072	Bacillus
31	837.2	37.0	2256	1	AB020073	Bacillus
32	837.2	37.0	2256	1	AB020074	Bacillus
33	837.2	37.0	2256	1	AB020075	Bacillus
34	837.2	37.0	2256	1	AB020076	Bacillus
35	837.2	37.0	2256	1	AB020077	Bacillus
36	837.2	37.0	2256	1	AB020079	Bacillus
37	837.2	37.0	2256	1	AB020080	Bacillus
38	837.2	37.0	2256	1	AB020081	Bacillus
39	837.2	37.0	2256	1	AB020082	Bacillus
40	837.2	37.0	2256	1	AB020083	Bacillus
41	837.2	37.0	2256	1	AB020084	Bacillus
42	837.2	37.0	2256	1	AB020085	Bacillus
43	837.2	37.0	2256	1	AB020086	Bacillus
44	837.2	37.0	2256	1	AB020087	Bacillus
45	837.2	37.0	2256	1	AB020088	Bacillus

ALIGNMENTS

RESULT 1
AR098263
LOCUS AR098263 2262 bp DNA linear PAT 14-FEB-2001
DEFINITION Sequence 5 from patent US 6074860.
ACCESSION AR098263
VERSION AR098263.1 GI:12807520
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2262)
AUTHORS Robertson,D.E., Sanyal,I. and Adhikary,R.S.
TITLE Catalases
JOURNAL Patent: US 6074860-A 5 13-JUN-2000;
FEATURES Location/Qualifiers

	source	1. .2262 /organism="unknown"	BASE COUNT ORIGIN	553 a	612 c	663 g	434 t		
	Query Match	100.0%; Score 2262; DB 6; Length 2262;							
	Best Local Similarity	100.0%; Pred. No. 0;							
	Matches 2262; Conservative	0; Mismatches 0; Indels 0; Gaps 0;							
Qy	1	ATGAATAACGCATCCGTGACGACTACACAGTACCTTGCGAGCAAAAGCATGCAGAGCATTT	60						
Db	1	ATGAATACGGATCCGCTGACGATCTACACAGTAGCTTGCGAGCAAAAGCATGCAGAGCATTT	60						
Qy	61	GTTCCCTTGSTATGCCAAGSCATAGAGCAATAAGGGAGAGAGCTATGAGCGGTAAATGT	120						
Db	61	GTTCCCTTGSTATGCCAAGSCATAGAGCAATAAGGGAGAGAGCTATGAGCGGTAAATGT	120						
Qy	121	CCTGTCATGACGGTGTATACACCTGACCGGTACTTCCACAAGAATTGGTGGCCGGAA	180						
Db	121	CCTGTCATGACGGTGTATACACCTGACCGGTACTTCCACAAGAATTGGTGGCCGGAA	180						
Qy	181	GGGTGAACCTGGATATTTTGCAATCAGCAAGATCGCAAAATCAGACCAGTGGATCCGGAT	240						
Db	181	GGGTGAACCTGGATATTTTGCAATCAGCAAGATCGCAAAATCAGACCAGTGGATCCGGAT	240						
Qy	241	TTCAACTACCGTGAAGAGTAGCCAAGCTCGATTTCCAGCGCTGAAGAAGATGTCCAC	300						
Db	241	TTCAACTACCGTGAAGAGTAGCCAAGCTCGATTTCCAGCGCTGAAGAAGATGTCCAC	300						
Qy	301	CGGTGATGACCGATAGCAAGAGTGTGGCCGCTGACTGGGGGCACTACGGCGTTTG	360						
Db	301	CGGTGATGACCGATAGCAAGAGTGTGGCCGCTGACTGGGGGCACTACGGCGTTTG	360						
Qy	361	ATGATCCGTATGGCTTGGCACTCCGCTGGCACCTTACCCTATTTGCTGATGCCGTGGGGC	420						
Db	361	ATGATCCGTATGGCTTGGCACTCCGCTGGCACCTTACCCTATTTGCTGATGCCGTGGGGC	420						
Qy	421	GGTGGTACCGGAAGCCAGCGCTTTGCAACGCTCAACTCCCTGGCCGGACAACGTCAGCGTG	480						
Db	421	GGTGGTACCGGAAGCCAGCGCTTTGCAACGCTCAACTCCCTGGCCGGACAACGTCAGCGTG	480						
Qy	481	GATAAAGCGCGCTCTGCTGTGGCCCATCAAGNAGAAAGTACGGCAACAANAATCAGCTGG	540						
Db	481	GATAAAGCGCGCTCTGCTGTGGCCCATCAAGNAGAAAGTACGGCAACAANAATCAGCTGG	540						
Qy	541	GCAGACCTGATGATTCCTGGCTGGCACCGTGGCTTATGAGTCCATGGGCTTACCTGCTTAC	600						
Db	541	GCAGACCTGATGATTCCTGGCTGGCACCGTGGCTTATGAGTCCATGGGCTTACCTGCTTAC	600						
Qy	601	GGCTTCCTTTTGGCCCGCTCGATATTTGGGAACCCGAAAAGATATCTACTGGSGTGAC	660						
Db	601	GGCTTCCTTTTGGCCCGCTCGATATTTGGGAACCCGAAAAGATATCTACTGGSGTGAC	660						
Qy	661	GAAAAAGAGTGGTGGCACTTCTGAGCAACGCTACGGCGACGTGTAACAAGCCAGAGACC	720						
Db	661	GAAAAAGAGTGGTGGCACTTCTGAGCAACGCTACGGCGACGTGTAACAAGCCAGAGACC	720						
Qy	721	ATGGAAAAACCGCTGGCGGCTGTCCAAATGGTCTGATCTATGTGAACCCGGAAGGTGT	780						
Db	721	ATGGAAAAACCGCTGGCGGCTGTCCAAATGGTCTGATCTATGTGAACCCGGAAGGTGT	780						
Qy	781	AACGGCCACCTGATCCGCTGAGAACCGCACAGAGTACTTGAACCTTTCCGCCGTATG	840						
Db	781	AACGGCCACCTGATCCGCTGAGAACCGCACAGAGTACTTGAACCTTTCCGCCGTATG	840						
Qy	841	CGGATGAACAGCAAAAAAACCGGACCGCTCACAGCTGGCGGCGCACACCGTCGGTAATGT	900						
Db	841	CGGATGAACAGCAAAAAAACCGGACCGCTCACAGCTGGCGGCGCACACCGTCGGTAATGT	900						
Qy	901	CACGGTAATGCAATGCTCTCGGTTAGCCCTGACCCCAAAAGCCCTGACGTTGAAAC	960						
Db	901	CACGGTAATGCAATGCTCTCGGTTAGCCCTGACCCCAAAAGCCCTGACGTTGAAAC	960						

||||| 2041 TGGAGCCGGTAGGACGCGCTACGAAATCCGGACCCAGACCGGTGCGGTGAAG 2100
QY 2101 TGGACCGCCTCGCGGTGGATCTGGTATTTGTTTCCAACTCGCTACTCGCTCTTTACGCA 2160
Db 2101 TGGACCGCCTCGCGGTGGATCTGGTATTTGTTTCCAACTCGCTACTCGCTCTTTACGCA 2160
QY 2161 GAAGTGACGCCAGGAGGAGTAACGGCGGAGAAAGTTTCGTCAGAGACTTCGTCGCGCCCTGG 2220
Db 2161 GAAGTGACGCCAGGAGGAGTAACGGCGGAGAAAGTTTCGTCAGAGACTTCGTCGCGCCCTGG 2220
QY 2221 ACCAAAGTGATGAACGCCGACCGTTTTCGACGCTCGCGTGTAA 2262
Db 2221 ACCAAAGTGATGAACGCCGACCGTTTTCGACGCTCGCGTGTAA 2262

RESULT 2
LOCUS AR215251 2262 bp mRNA linear PAT 25-SEP-2002
DEFINITION Sequence 5 from patent US 6410290.
ACCESSION AR215251
VERSION AR215251.1 GI:23313418
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2262)
AUTHORS Robertson,D.E., Sanyal,I. and Adhikary,R.S.
TITLE Catalases
JOURNAL Patent: US 6410290-A 5 25-JUN-2002;
FEATURES Location/Qualifiers
source 1..2262 /organism="unknown"
BASE COUNT 553 a 612 c 663 g 434 t

Query Match 100.0%; Score 2262; DB 6; Length 2262;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAATACGCATCCGCTGACGATCTACACAGTAGCTTGCAGCAAGATGACAGACATTT 60
Db 1 ATGAATACGCATCCGCTGACGATCTACACAGTAGCTTGCAGCAAGATGACAGACATTT 60
QY 61 GTTCCCTTGGTATGCCAAGGCATAGACCAATAAGGAGAGAGCTATGAGCGGTAATGT 120
Db 61 GTTCCCTTGGTATGCCAAGGCATAGACCAATAAGGAGAGAGCTATGAGCGGTAATGT 120
QY 121 CCTGTCATGCAGGTGGTAAACACCTCGACCGGTACTTCCAAAGATTTGGTGGCGGAA 180
Db 121 CCTGTCATGCAGGTGGTAAACACCTCGACCGGTACTTCCAAAGATTTGGTGGCGGAA 180
QY 181 GGGTTGAACCTGGATATTTTCATCAGCAAGATCGCAATCAGACCCGATGGATCCGGAT 240
Db 181 GGGTTGAACCTGGATATTTTCATCAGCAAGATCGCAATCAGACCCGATGGATCCGGAT 240
QY 241 TTCAACTACCGTGAAGAAGTAGGACGATCGATTCGACGGCTGGAAGAAAGATGTCAC 300
Db 241 TTCAACTACCGTGAAGAAGTAGGACGATCGATTCGACGGCTGGAAGAAAGATGTCAC 300
QY 301 CGCTTGATGACCGATAGCAAGAGTGGTGGCCGCTGACTGGGGCACTACGGCGTTTG 360
Db 301 CGCTTGATGACCGATAGCAAGAGTGGTGGCCGCTGACTGGGGCACTACGGCGTTTG 360
QY 361 ATGATCCGTATGGCTTGGCACTCCGCTGACCTACCGTATTTGCTGATGGCGTGGGGC 420
Db 361 ATGATCCGTATGGCTTGGCACTCCGCTGACCTACCGTATTTGCTGATGGCGTGGGGC 420
QY 421 GGTGGTACCGAAGCCAGCGCTTTGCAACCGCTCAACTCTCGCCGGACAAAGTCAGCGCTG 480
Db 421 GGTGGTACCGAAGCCAGCGCTTTGCAACCGCTCAACTCTCGCCGGACAAAGTCAGCGCTG 480
QY 481 GATTAAGCGCCGCTCTCTCTGTGCCGNTCAAGAGAAGTACGCAACAAATCAGCTGG 540

Db 481 GATTAAGCGCCGCTCTCTCTGTGCCGNTCAAGAGAAGTACGCAACAAATCAGCTGG 540
QY 541 GCAGACCTGATGATTTCTGGCTGGCACCGCTGGCTTATGATCCATGGCTTACCTGCTTAC 600
Db 541 GCAGACCTGATGATTTCTGGCTGGCACCGCTGGCTTATGATCCATGGCTTACCTGCTTAC 600
QY 601 GGCTTCTTTTCGGCCGCTCGATATTTGGGAACCCGAAAGAGATATCTACTGGGGTGAC 660
Db 601 GGCTTCTTTTCGGCCGCTCGATATTTGGGAACCCGAAAGAGATATCTACTGGGGTGAC 660
QY 661 GAAAAGAGTGGCTGGCACCTTCTGACGAACCGCTACGGGGAGGTGAACAGCCAGAGACC 720
Db 661 GAAAAGAGTGGCTGGCACCTTCTGACGAACCGCTACGGGGAGGTGAACAGCCAGAGACC 720
QY 721 ATGAAAAACCCGCTGGCGGCTGTCCAAATGGTCTGTATGTATGTCAACCCGGAAGTGTT 780
Db 721 ATGAAAAACCCGCTGGCGGCTGTCCAAATGGTCTGTATGTATGTCAACCCGGAAGTGTT 780
QY 781 AACGGCCACCTGATCTCGCTGAGAACCCGACAGCAGTACTTGAACCTTCGCCCGTATG 840
Db 781 AACGGCCACCTGATCTCGCTGAGAACCCGACAGCAGTACTTGAACCTTCGCCCGTATG 840
QY 841 GCGATGAACGACGAAAAAACCCGACGCTCAGACGTGGGGCCACACCGTCGGTAATGT 900
Db 841 GCGATGAACGACGAAAAAACCCGACGCTCAGACGTGGGGCCACACCGTCGGTAATGT 900
QY 901 CACGGTAATGCAATGCTCTGCGPTAGCCCTGACCCAAAGACCTCTGACGTTTCAAAAC 960
Db 901 CACGGTAATGCAATGCTCTGCGPTAGCCCTGACCCAAAGACCTCTGACGTTTCAAAAC 960
QY 961 CAGGCTTTAGGTTGGGGCAACCCCAACATGCAAGGCAAGGCAAGCCGTCACCTCG 1020
Db 961 CAGGCTTTAGGTTGGGGCAACCCCAACATGCAAGGCAAGGCAAGCCGTCACCTCG 1020
QY 1021 GGTATCGAAGTGTGTTGACCAACCAACCCGACGAAATTCGATATGGGCTATTTGACCTG 1080
Db 1021 GGTATCGAAGTGTGTTGACCAACCAACCCGACGAAATTCGATATGGGCTATTTGACCTG 1080
QY 1081 CTGTTTCGGCTCAATTTGGGAACCTGAAAAAGAGTCTCTCGCGTGGCCACCATTTGGGAACCG 1140
Db 1081 CTGTTTCGGCTCAATTTGGGAACCTGAAAAAGAGTCTCTCGCGTGGCCACCATTTGGGAACCG 1140
QY 1141 ATTGACATCAAAAAGGAAAAACAGCCGTTGACGCCAGCAGCCCTCTATTTCGCCACAAAC 1200
Db 1141 ATTGACATCAAAAAGGAAAAACAGCCGTTGACGCCAGCAGCCCTCTATTTCGCCACAAAC 1200
QY 1201 CCGATCATGACCGATGCGGATATGCGGATTAAGGTAAATCCGACCTATCGGCTATCTGC 1260
Db 1201 CCGATCATGACCGATGCGGATATGCGGATTAAGGTAAATCCGACCTATCTGCCTATCTGC 1260
QY 1261 GAAAAATTCATGGCCGATCTCTGAGTACTTTCAAGAAAACTTTTCGCGAAGCGCTGGTTCAAG 1320
Db 1261 GAAAAATTCATGGCCGATCTCTGAGTACTTTCAAGAAAACTTTTCGCGAAGCGCTGGTTCAAG 1320
QY 1321 CTGAGCAGCCGTGACCTGGGCCCCGAAATACAGTTTACATCGGGCCCCGGAAGTGGCCGAGAA 1380
Db 1321 CTGAGCAGCCGTGACCTGGGCCCCGAAATACAGTTTACATCGGGCCCCGGAAGTGGCCGAGAA 1380
QY 1381 GACCTGATTTGCGAAGACCCGNTTCCGCGAGGTAAACCGACTACTGCGAAGAGTGGTCTC 1440
Db 1381 GACCTGATTTGCGAAGACCCGNTTCCGCGAGGTAAACCGACTACTGCGAAGAGTGGTCTC 1440
QY 1441 AAGCAGAAAAATTCACAAAGTGGCCTGAGCATTTAGTGAGATGGTCTCCACCGCTGGGAC 1500
Db 1441 AAGCAGAAAAATTCACAAAGTGGCCTGAGCATTTAGTGAGATGGTCTCCACCGCTGGGAC 1500
QY 1501 AGTGCCCGTACTTATCGCGGTTCCGATATGCGCGCGGTGCTTAACGGTCCCGCATTCGC 1560
Db 1501 AGTGCCCGTACTTATCGCGGTTCCGATATGCGCGCGGTGCTTAACGGTCCCGCATTCGC 1560
QY 1561 TTGGCCCCACAGAAGCAGTGGCAGGGCAACGAGCCGAGCGCTGGCGGAAGTGTGCTGAGC 1620
Db 1561 TTGGCCCCACAGAAGCAGTGGCAGGGCAACGAGCCGAGCGCTGGCGGAAGTGTGCTGAGC 1620

```
Db 1561 TTGCCCCCAGAACGAGTGGCAGGCGCAACGAGCGCGGAGCGCTGGCGGAAGTGTGTCGAGC 1620
QY 1621 GTCTACGACGAGATCTCTGCCGACACCGCGCGTAGCATCGCGGACGTGATCGTTCTGCGCC 1680
Db 1621 GTCTACGACGAGATCTCTGCCGACACCGCGCGTAGCATCGCGGACGTGATCGTTCTGCGCC 1680
QY 1681 GGTAGCGTAGGCATCGAAGAACCGCGGAAGACGACGAGTTACGATGTGCGCGTCCCTTC 1740
Db 1681 GGTAGCGTAGGCATCGAAGAACCGCGGAAGACGACGAGTTACGATGTGCGCGTCCCTTC 1740
QY 1741 CTGAAGAGCGCGTGGCGATGGCGACCGCGGAGATGACGACGACGACTCTTCGCGACCGCTG 1800
Db 1741 CTGAAGAGCGCGTGGCGATGGCGACCGCGGAGATGACGACGACGACTCTTCGCGACCGCTG 1800
QY 1801 GAGCCGCTGCGCGATGGCTTCCGCAACTGGCAGAAAGAGTATGTGTGAAGCCGCGAA 1860
Db 1801 GAGCCGCTGCGCGATGGCTTCCGCAACTGGCAGAAAGAGTATGTGTGAAGCCGCGAA 1860
QY 1861 GAGATGCTGTGATGCTGGCGACGCTGATGGGCTTAACCGGCGCGGAATGACCGGTGCTG 1920
Db 1861 GAGATGCTGTGATGCTGGCGACGCTGATGGGCTTAACCGGCGCGGAATGACCGGTGCTG 1920
QY 1921 CTGGCGGTATGCGCTGACGACCAACTATGTTGGTGGCAGCAACACGCGCGTATTCACC 1980
Db 1921 CTGGCGGTATGCGCTGACGACCAACTATGTTGGTGGCAGCAACACGCGCGTATTCACC 1980
QY 1981 GATTGTGAAGCCGAGTTGACCAACGACTTTTTTGTGAACCTGACCGATGCGGGAACAGC 2040
Db 1981 GATTGTGAAGCCGAGTTGACCAACGACTTTTTTGTGAACCTGACCGATGCGGGAACAGC 2040
QY 2041 TGAAGCCGCTAGGTAGCAACGCTAGCAATCCGCGACCGCAAGACCGGTGCGGTGAAG 2100
Db 2041 TGAAGCCGCTAGGTAGCAACGCTAGCAATCCGCGACCGCAAGACCGGTGCGGTGAAG 2100
QY 2101 TGGACCGCTCGCGGTGGATCTGTTGTTTCCAACTCGCTACTGCGCTCTTAGCA 2160
Db 2101 TGGACCGCTCGCGGTGGATCTGTTGTTTCCAACTCGCTACTGCGCTCTTAGCA 2160
QY 2161 GAAGTGTACGCCAGGACGATACCGCGAAGTCTCGTCAGAGACTTCGTCGCGCGCTGG 2220
Db 2161 GAAGTGTACGCCAGGACGATACCGCGAAGTCTCGTCAGAGACTTCGTCGCGCGCTGG 2220
QY 2221 ACCAAGTGTAGTAAACCGCGCGCTTCGAGCTGCGCTCGTAA 2262
Db 2221 ACCAAGTGTAGTAAACCGCGCGCTTCGAGCTGCGCTCGTAA 2262

RESULT 3
AE004233
LOCUS
DEFINITION
Vibrio cholerae chromosome I, section 141 of 251 of the complete
chromosome.
ACCESSION
AE004233 AE003852
VERSION
AE004233.1 GI:9656055
KEYWORDS
Vibrio cholerae
SOURCE
Vibrio cholerae
ORGANISM
Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
Vibrionaceae; Vibrio.
1 (bases 1 to 14482)
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A.,
Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I.,
Sellers, P., McDonald, L., Utterback, T., Smith, H.O., Colwell, R.R.,
Nierman, W.C. and White, O.
TITLE
DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae
JOURNAL
Nature 406 (6795), 477-483 (2000)
MEDLINE
20406833
PUBMED
10952301
REFERENCE
2 (bases 1 to 14482)
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,
```

```

TITLE
JOURNAL
FEATURES
source
1..14482
/organism="Vibrio cholerae"
/mol_type="genomic DNA"
/strain="N16961"
/serotype="O1"
/db_xref="taxon:666"
/chromosome="I"
/notes="biotype: El Tor"
190..1494
/gene="VC1549"
190..1494
/gene="VC1549"
/notes="similar to SP:PI0904 GB:U00039 GB:M33735 PID:43244
PID:466589; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="glycerol-3-phosphate ABC transporter,
periplasmic glycerol-3-phosphate-binding protein"
/protein_id="AAF94703.1"
/db_xref="GI:9656056"
/transl_table=11
/translation="MNKVMLTGLLTATAGSVFASASTLAATEITMWHAMGGQGETVYQ
LATDFNASQDYKLTVPYKSYTETLTAGIAFRAQAPNLIQVFDAGATIMSSKGV
AKPVQDLMTSGYPTAQDYIAGRVNFYADNTGKVMGFNSRNLPTVLYYKMDLASVN
AOPATYEOLEQVAALAEKGAQFASLTPTWIMFENFNSRNLPLSDKQNGQDLST
KINFSDMLMHVKLKESDGYKYGSDWANTQTPERGEVAMHMGSSGSGFGLR
QMNPLNCTTLYPWSVTNPNHTTIGGAALFALGHTPAQDKGVAATFAYLTKEPT
QNMHKTGTVPVNTAAIDLTQSGYIQENPDVAGVKGOLSLPDGEMTKGYRIGYTPQ
VREVMHREFDNIFSGRSTVESSLNKVEDESAKLLNRFATVQ"
1570..2451
/gene="VC1550"
1570..2451
/gene="VC1550"
/notes="similar to SP:PI0905 GB:U00039 PID:43247 PID:466588
GB:U00096; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="glycerol-3-phosphate ABC transporter, permease
protein"
/protein_id="AAF94704.1"
/db_xref="GI:9656057"
/transl_table=11
/translation="MQRQHFHTPLPYLLAPQIVIAVFFIYPAQAQIYLSFMLED
PMGLSFTFVWFENYQVLFSDYLDLSGFTLLFSLVSVFLSLALALLAVKADNIHNG
QGSYRVTLTWYAVAPAVAGLGGFLFNPHIGVLTDFQYIGWQSFQDPVDATIAL
ILVSVWKQSVNPLYFLAGLSISYAVREAAALDCQSDSKRFSWITFLLAPDTGFFLL
VINLYTAFEDTFGVIDTMTNGGPGGTTSLVYKVRDGFVGADLGGSAQSVVLLVLY
LALTWTQFRFEKRVHY"
2461..3303
/gene="VC1551"
2461..3303
/gene="VC1551"
/notes="similar to SP:PI0906 GB:U00039 PID:43248 PID:466587
GB:U00096; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="glycerol-3-phosphate ABC transporter, permease
protein"
/protein_id="AAF94705.1"
/db_xref="GI:9656058"
/transl_table=11
/translation="MKSNLSDHLIILAGALFMLVPLMLIFASSTHPTNTIYSEGLQW
TLGDNLISAVYREAWNKSGLCFAGNVTAGTWTVMNSMGLGFAIGKIVISMAAYALYF
RLPYASAWFWLIFTILLPLEVRIIPSYEVVSNLGMNSYTGVLVPLIASATATFFFR
QFKFTIPDELLAAQDNLNAGPIRFIDILFPLSKTMMAAFIIMFVVGWQNYLWPLIML
TTDEQNTIVMGIKQILNNINETSIPRYDYAFAMVILAMLPVLVYVYIFQRFVFKGLV
```

Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Utterback, T., Fleishmann, R.D., Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R., Mekalanos, J.J., Venter, J.C. and Fraser, C.M.
Direct Submission
Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA

Location/Qualifiers


```

ESEK"
3293.. .4474
/gene="VC1552"
CDS
/gene="VC1552"
/note="similar to SP:P10907 GB:U00039 PID:43249 PID:912455
GB:U00096; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="glycerol-3-phosphate ABC transporter,
ATP-binding protein"
/protein_id="AAF94706.1"
/db_xref="GI:9656059"
/translation="MKNNQSNHLLRSKHAEPVMDIKOLVTTYNGHQAVKGVDL
SIHGEEIVLVGPGCKSSILRSIAGLESITGEIHLAGRVNDEKPNARDIAVFQ
NYALYPMVYDNLAYGLKNNRVDKQTAIAKIAKVAIKLIEEYLDKRPKLSCGQRQ
RVAMGRAIVRDPQFLFDEPLSNLDAALRAHMLRIKTLQRELGVTSVYVTHDOVEAM
TLADRIYVVGGEIQGTAEVYHOPASTVAFSIGSPAMNELPASIKOGSLHIAGK
HCYLPQDMASSENITGIRPEHASLHPLNAIELKLDIQVEPLGNOLVHGKIIGL
ESEDFTAVTAIEIDVHQTLPIWVALEQLHFDQGRFVQSHRSCTQSSKVASTRQ
Q"
4499.. .5260
/gene="VC1553"
CDS
4499.. .5260
/gene="VC1553"
/note="similar to SP:P09392 GB:M96795 PID:146186
PID:606358 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="glycerol-3-phosphate regulon repressor"
/protein_id="AAF94707.1"
/db_xref="GI:9656060"
/translation="MTLLSHRQKQILEYLOQTHGVLSALLSERFEVSQVIRKDMND
LSDKGMVRRVHGGLSPNTNLSFNRLNINLAKORIAORVAQELPEGCSIFLGIG
TTPQVAQLIDHPGLVTVTNNLAALTLSRNPINLHLAGLVHRHSDERVGEATRR
EYRSFNKVGIFGVLNSRGQLDFTPEANLIRAEIHEHSEQLWLVADKSLERYAP
VYSELSEMRHLFVDKSTPAFQQLSLIYHVELIES"
5288.. .6001
/gene="VC1554"
CDS
5288.. .6001
/gene="VC1554"
/note="similar to SP:P54527 PID:1303930 GB:AL009126;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="glycerophosphoryl diester phosphodiesterase,
putative"
/protein_id="AAF94708.1"
/db_xref="GI:9656061"
/translation="MMPSSIIVHRGVAGTYPTNTRVSVQAAINLGLSWVEDVDQPTKD
NLVYVCHDHTIDRCNSGKRIDEYTLAEKQLDFGAWDFPRFAGETIMTLBELLVLA
EHLGLNIEVXVDKHVASVVOQLKAQDQPLVADKTLILSSFSHDILQLHQCPCGY
KLGILSERLTORDRQVLKQVNAFSCNLYRWITEKHIRDLANGGYQIWCYTLNQDKF
RYLDKVAIFSDPFERFIN"
complement(6077.. .6589)
/gene="VC1555"
CDS
complement(6077.. .6589)
/gene="VC1555"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF94709.1"
/db_xref="GI:9656062"
/translation="WRLRASSINKETPMFIGHTQRFCAALSPOLQRLINEVLQRFV
AAPLPTGHEIQGSAFLVMEHDTQPLARRSECHARYLDVQILLOQRERFGYSLAP
FSGLEDLADTARVAFSAQLVEERFVLAAGDFIVFPGQPHRLIAVGEGBEPVRKY
VIKVDKAFPE"
complement(6558.. .6920)
/gene="VC1556"
CDS
complement(6558.. .6920)
/gene="VC1556"

/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF94710.1"
/db_xref="GI:9656063"
/translation="WTHIRNPGRFMSDITVFNGIAHFVEVADSDTSADMGOQVEQVL
AQAEQGLAKIGSDRSRLISVTVYTFDFADLPVLIQINDAWPEGCAPSRACVKAELAD
PDYKEMAFVAASGEQYO"
complement(7055.. .8089)
/gene="VC1557"
CDS
complement(7055.. .8089)
/gene="VC1557"

/note="similar to PID:892607 GB:U00096 PID:1789068;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, LacI family"
/protein_id="AAF94711.1"
/db_xref="GI:9656064"
/translation="MRTQFEGLLMATINDVCKLAGVSKATVSRVLNETGOVKAOTREA
VLAAMQOLGYQPNLSLAQALNTNTNSIGLVLPHPSSYFSGILPEAEOGAKGKLL
VMNSKNSQEGKEAVATLAAORCDAILLYSRLNEVQLLELQOQHPSPLYILNRLHH
POLHSFGLDQRIAGLAQHMLNHLHROICITSPVSETKIRIQVYQQAHLQEGIE
LSSLSFIEGDMTLLGGYQAMQOLLQGGISMTAVFACNDDMALGRAMHEHGIHVPKQ
VSIIGIDNEPAAFAIPSLSSVSLPIGELTQQAISLAVEIANKKPQDAQHRLYQSLI
ARESTIALKL"
complement(8125.. .9561)
/gene="VC1558"
CDS
complement(8125.. .9561)

Query Match 50.8%; Score 1149; DB 1; Length 14482;
Best Local Similarity 70.9%; Pred. No. 3.6e-230;
Matches 1524; Conservative 0; Mismatches 625; Indels 0; Gaps 0;

QY 109 AGCGGTAATGCTGTCATGCACGGTGGTACACCTGCACGGGTACTTCCACAAAGATC 168
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 998 AGCGGTAATGCCCTGTCATGCACGGTGGTACCTGCACGGAGCATGTCCACCATGGAC 10057
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 169 TGGTGGCGGGAAGGGTGAACCTGGATATTTTGCATCAGCAAGATCGCAAAATCAGACCGG 228
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 10058 TGGTGGCGCTAAAGCGCTAACTTAGACATCTTCATCAGCATGACAGCAAAACCAATCCA 10117

QY 229 ATGGATCCGGATTTCAACTACCGTGAAGAAGTACGCAAGCTGATTTTCACGCGCTGAAG 288
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 10118 CTTGCTGCTGATTTCAACTATCGCGAAGAGCTGAAAAAGCTTGATGTAGAACGCTAAAA 10177

QY 289 AAAGATGTCACGCGCTTGTATGACCATAGCCAAAGAGTGGTGGCCGCTGACTGGGGGAC 348
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 10178 CGCGATCTGAAAGCGCTTAATGACCAATAGCCAAAGAGTGGTGGCCGCGGACACTGGGGACAT 10237

QY 349 TACGCGCGTTTTCATCATCGGTATGGCTTGGCAGCTCCGCTGGCAGCTACCGTATTCGTGAT 408
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 10238 TACGCGCGTTTTCATCATCGGTATGGCTTGGCAGCTCTGGGGTACCTACCGTATTCGCCGAT 10297

QY 409 GCGCTGGGGCGGTGGTACCGAAGCCAGCGCTTTGCACCGCTCAACTCTCTGGCGGAC 468
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 10298 GGTGCTGGCGGCGGTACCGTAAACCAACGTTTGGCCCACTGAATCATGCGCTGAT 10357

QY 469 AACGTCAGCTGGATAAAGCGCGCTGCTGCTGGCGCATCAAGAAGATGACGGCAAC 528
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 10358 AACGCCAACTTAGATAAAGCACGCGCTTGTGTCGCTATCAACAAAGATGACGGTAC 10417

QY 529 AAAATCAGCTGGCAGACCTGATGATCTGGCTGGCAGCGCTGCTGATGATGATGATGATGATGAT 588
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 10418 AAAATCAGCTGGCGGATTTGATGATCTTCGCGGGAACATGCGTATGATGATGATGATGATGAT 10477

QY 589 TTACCTGCTTACGGCTTCTCTTCGCGCGCGCTGCGATATTTGGGAACCCGCAAAAGATATC 648
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 10478 CTCAAAACCTTTGGTTTTCCTTTGGCGCGGAAGACATCTGCGACCCGCAAAAGACATT 10537

QY 649 TACTGGGTGACGAAAAAGAGTGGCTGGCACCTCTGACGAAACGCTACGGCGGACGCTGAAC 708
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Ds	10538	TACTGGGGTTCTGAAAAAGAGTGGCTTGGCCAAATCTGGCGGTGAAAAATAGCCGTACTCT	10597
Qy	709	AAGCCAGAGACCAATGGAACCCCGCTGGCGGCTCTCCAAATGGGTCTGATCTATGTGAAC	768
Ds	10598	GGTCAGCGCATCTCGAAAAACCCACTTGGCGCGTCATGATGGGCTTATCTAGTCAAC	10657
Qy	769	CCGGAAGGTGTTAACGGCCACCTTGATCCGCTGAGAACCGCACAGCAGGTACTTTGAACCC	828
Ds	10658	CCAGAAGGTGTGATGTAACCCAGATCCACTCAAAACGGCGCAAGACATGCGTGTCACT	10717
Qy	829	TTGCCCGTATGGCATGAACGAGCAAGAAACCCAGCCCTCAGCTGAGCGTGGGCGCACACC	888
Ds	10718	TTTGCTCGCATGGCAATGAATGAGAAAGACCGTTGCTGAGCGGTGGCGGCATACG	10777
Qy	889	GTCGGTAAATGTCAACGGTAATGGCAATGCTCTGCTTAGCCCTGACCCCAAAAGCCCTCT	948
Ds	10778	GTGGTAAAGCGCACGCTAATGGCAAGCTTCTAACCTTGGCCCTGATCCTGAAGGCGCT	10837
Qy	949	GAGCTTGAACACAGGCTTAGTTGGGGCAACCCCAACATGCAAGGCAAGGCAAGCAAC	1008
Ds	10838	GAACTGCACGAACAAGGCTTAGTTGGAAACAACACACACAGCCGTGGGATTGGTGCAGAA	10897
Qy	1009	CGCTGACCTCGGTATTCGAAGTGTGGACACCAACCCACAGAAATTCGATATGGC	1068
Ds	10898	ACCGTACCAAGTGTATGAAGTGTGATGACCACTCACCACCCCTTGGGATAAGCA	10957
Qy	1069	TATTCGACCTGCTGTTCGCTACAAATTTGGGAACATGAAAAGAGTCTCGCGGTGCCAC	1128
Ds	10958	TATTTCTACTTACTGCTCAGCTATGAATGCGAGCTCACTAAGACCCCTGCAAGTGCATG	11017
Qy	1129	CATTTGGAACCGATTGACATCAAAAGGAACAAAGCCGTTGACGCCAGCGACCCCTCT	1188
Ds	11018	CAATGGGAACCTGTCAACATCAAGAGGAAGATAAACCCGTTGATGTGGGAAGACCCAAAT	11077
Qy	1189	ATTGCGCACACCGCATCATGACCGATGCGGATGCGGATGAAGTAAATTCGACCTAT	1248
Ds	11078	ATTGCTACAAACCAATGATGACCGATGCGGATGCGGATGCGGATGCGGATGCGGAT	11137
Qy	1249	CGCGCTATCTCGCAAAATTCATGGCGATCTCGGATCTCTGAGTACTTCAAGAAACCTTTGCGGAAG	1308
Ds	11138	CGAAGATTTTCAGAGCCCTTCTACAAAGATCCCGCTTACTTCTCAGAAGTGTTCGCGGA	11197
Qy	1309	CGGTGTTCAAGCTGACGACCGTGTACCTGGCGCCGGAATCAGTTTACATCGGCCCGGAA	1368
Ds	11198	CGTGTGTTAAGTCTACTCAGCGCATGCGGATGCGGATGCGGATGCGGATGCGGAT	11257
Qy	1369	GTGCGCGCAGAACCTGATTTGGCAAGCCGATTCGCGCAGTACACCGACTACTGC	1428
Ds	11258	GTTCGCGCAGAAATTTGATCTGCGCAGACCCCTGTTCCAGCAGTCTGTAAGATATGAT	11317
Qy	1429	GAAGAAGTGTCAAGCAGAAATTTGCAAAAGTGGCTGAGCATTTAGTGAGATGGTCTCC	1488
Ds	11318	GTCAATCGGTCAAAAGCAAAATTCGCGCAAGCGGTCTAAGCATCACTGAATGTTGTCAGC	11377
Qy	1489	ACCGCTTGGACAGTGGCTTACTTATCGCGGTTCGATATGCGCGCGGTGCTAACGGT	1548
Ds	11378	ACCGCTTGGATAGCGTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	11437
Qy	1549	CGCGCATTCGCTTGGCGCCACAGAGAGTGTGGCGGCAACGAGCGCGGCGCTGCGC	1608
Ds	11438	CGCGGTATTCGCTCTGCTCAAAAGATTTGGAAGCAAGCACTGCGGCTGCTGGT	11497
Qy	1609	AAAGTGTGAGCGTCTACGACAGATCTCTGCGGACACCGCGCTAGCATCGCGGAGGTG	1668
Ds	11498	AAAGTCTGCGGTATTAGAGAAATTTGCGGCTGAACTCGCATCACTAGTACGCGATACG	11557
Qy	1669	ATGCTTTCGCGGTAGCTAGCATCGAAGAAACCCGCAAGACGAGCGATTACGATGTG	1728
Ds	11558	ATTGTTCTGCTGTACAGTATGATGAAACAAGCGGCTTAAAGCGCGGCGTGAATGTG	11617
Qy	1729	CGGTTTCCTTCTGAAAGCGGTGGCGATGCGACCCCGCATGATGACGACGAGCTCC	1788
Ds	11618	ACCGTACCGTTCGACCGGCGGCGATGCGACAAATCGCAACAAACCGGATGTTAGTCC	11677
Qy	1789	TTGCACCGCTGGAGCGCTGGCGATGGCTTCCGCAACTTCCGCAAGAAAGAGTATGTG	1848
Ds	11678	TTTGAAGTGTGAGCTCTCGCGATGGCTTCCGTAAGTGGCAAGAAAGCATATGTG	11737
Qy	1849	GTGAAGCCGGAAGAGATGCTGCTGGATCGTGGCGAGCTGATGGCTTAAACGGCCGGA	1908
Ds	11738	GTCAACCGGAAGAAATTTGCTCGAACAAGCGCACTGCTGCGCTCACCGCTCCAGAA	11797
Qy	1909	ATGACCGTGTGCTGGCGGCTATGCGGCTACTGGGCAACCACTATGTTGGCAACCAAC	1968
Ds	11798	ATGACGCTGTAATTTGTTGGCATGCTGCTGGCTACTAAGTACGTTGGCGGCAACAC	11857
Qy	1969	GGCTATTTCACCTTGTGAAGCGCATTTGACCAACGACTTTTGTGTAACCTGACCGAT	2028
Ds	11858	GGCGTGTTCACGACCGCTGGGTGCTTGACCAACGATTTCTTCGTTAACCTCAGCGAC	11917
Qy	2029	ATGGGGAACAGCTGGAAGCGGTAGGTAGCAACGCTTACGAAATTCGCGGACCGCAAGAC	2088
Ds	11918	ATGAGTTACACTTGGAAACCACTGTCGCAACTCTTACGAGATTTGTTGAACGTAAGAC	11977
Qy	2089	GGTCCCGTGAAGTGGACCGCTCGCGGTGATCTGCTGATTTGTTGTTTCAACTCGCTACTG	2148
Ds	11978	GGCAAGTGAAGTGGACGACCTGCTGTTGATCTGTTGTTGTTGTTTCAACTCGCTACTG	12037
Qy	2149	CGCTCTTACGCAAGTGTACGCGCAGGAGATTAACGCGGAGAAAGTTCGTCAGAGACTTC	2208
Ds	12038	CGTCTTATCGGAAGTGTACGCGACAGCAGCAACAAGAAATTCGTCAGAGACTTC	12097
Qy	2209	GTGCGCCCTGGACCAAGTGTATGACGCGGACCGCTTTCGACGTCGCGT 2257	
Ds	12098	GTTCGCGCATGGCAAGTAATGAATGCGGATCGATTTGATCTGCTGT 12146	
RESULT 4	AE008223/c		
LOCUS	AE008223	7772 bp	DNA linear BCT 18-DEC-2001
DEFINITION	Agrobacterium tumefaciens str. C58 linear chromosome, section 27 of 187 of the complete sequence.		
ACCESSION	AE008223	AE007870	
VERSION	AE008223.1	GI:15158557	
KEYWORDS	Agrobacterium tumefaciens str. C58 (Cereon)		
SOURCE	Agrobacterium tumefaciens str. C58 (Cereon)		
ORGANISM	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.		
REFERENCE	1 (bases 1 to 7772)		
AUTHORS	Hinkle, G., Slater, S.C. and Goodner, B.		
TITLE	Complete Genome Sequence of Agrobacterium tumefaciens C58 (Rhizobium radiobacter C58), the Causative Agent of Crown Gall Disease in Plants		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 7772)		
AUTHORS	Hinkle, G., Slater, S.C. and Goodner, B.		
TITLE	Direct Submission		
JOURNAL	Submitted (14-AUG-2001) Bioinformatics, Cereon Genomics, 45 Sidney Street, Cambridge, MA 02139, USA		
COMMENT	Approximately 800 bp of telomeric sequence missing from the left end of the chromosome and 200 bp missing from the right end.		
FEATURES	location/Qualifiers		
source	1..7772		
gene	/organism="Agrobacterium tumefaciens str. C58 (Cereon)"		
CDS	/mol_type="genomic DNA"		
	/strain="C58"		
	/db_xref="taxon:181661"		
	93..962		
	/gene="AGR_L_475"		
	93..962		
	/gene="AGR_L_475"		
	/note="(AF302797) PobR"		
	/codon_start=1		
	/transl_table=11		
	/product="AGR_L_475p"		

QY	532	ATCAGCTGGGCAGACACCTGATGATTTCTTGGCTTGGCACCGCTGGCTTATGAGTCCATATGGCGTTA	591
Db	3085	ATTTCCTGGGCCGACCTTATCGCGCTCGCCGGCACGATCGCTTATGAGTCGCGGGTCTG	3026
QY	592	CCTGGCTTACGGCTTCTCTTTCGGCCCGGCTCGATATTYTGGAAACCCGGAANAAGATATCTAC	651
Db	3025	AAGACCTTCGGCTTCGCGCTTCGCGCCGGAAGACATCTGGCGCGGAAAGGACACCTAT	2966
QY	652	TGGGGTGACAAAAGAGTGGCTGGCACCTTCTTCAGCAAGCTTACGGCAGCTGACGGACGTGAACAAG	711
Db	2965	TGGGGTGACAAAAGATGGCTTGGCCCGCAGCAGCGCGGTTATGGCAGCTGAGCAAG	2906
QY	712	CCAGAGACCATGAAAACCCGCTGGGGGCTGTCCAAATGGGTCTGATCTATGTGAAACCCG	771
Db	2905	CCCGAGACGCTGAAAACCCGCTTTCGCCGCTGCAGATGGGCTGATCTACGTCACACCCG	2846
QY	772	GAAGGTGTTAAACGGCCACCTGATTCGCTGAGAACCGCACAGCAGGTACTTTGAAACCTTC	831
Db	2845	GAAGGTGTCAACGGCAAGTCCGATCCGCTGGCGACGGCGCGCAGATGCGCGAAACCTTT	2786
QY	832	GCCCGTATGGCATGAACGACGAAAAAACCGCAGCCCTCAACGCTGGCGGCCACACCGTC	891
Db	2785	GCCCGCATGGGATGAGTACGAGGAAACGCTTGCCCTGACGGCCGGCGCCACACCATC	2726
QY	892	GSTAAATGTCACGGTAAATGGCAATGGCTTCGCTTAGCCCCCTGACCCAAAGCCTCTGAC	951
Db	2725	GGCAAGTCCCATGGCAATGGCAGTGTGTCCAATCTCAGCCCGATCCGGAAGCTGCTGC	2666
QY	952	GTTTAAACACGAGGCTTAGGTTGGGCGAACCCCAACATCGAGGGGCAAGGCACGACGCC	1011
Db	2665	CCGGAATACAGGGTCTCGGCTCGATCAATACCAAGGGCGCGCGCATTTGGCCGTGACACC	2606
QY	1012	GTGACCTCGGGTATCGAAGGTGCTTGGACACCAACCCACGAAATTCGATATGGGCTAT	1071
Db	2605	GTGGTGTGGGTATCGAGCGCATGGACAAAGCGAACCAAGTGGGACACAGGCTTC	2546
QY	1072	TTGCACTGTCTTTCGGCTACAATTTGGGAACGTGAAAGAGTCTCTCCGCTGCCACCAT	1131
Db	2545	TTGCACTGTCTTTCAGCACAGTGGACCTTGACGCACAGCCCGCGGTGCATCGCA	2486
QY	1132	TGGGAACCGATTGCATCAAAAGGAACACACCCCGTTGACGCCAGCAGCCCTCTATT	1191
Db	2485	TGGGCGCGATTACCATCGCCGGAAGAACACAGCCCTTGTGATCGAGGATGCGCTGATC	2426
QY	1192	CGSCACAACCCGATCATGACCGATCGCGATATGGCGATAAAGGTAAATCCGACCTATCGC	1251
Db	2425	CGCACCATCCGATGATGACGACCGGACATGGCCCTGAGGTCTGATCGATCTACCGC	2366
QY	1252	GCTATCTCGGAAAAATTCAATGGCCGATCTGTAGTACTTCAAGAAAACTTTCGCAAGGCG	1311
Db	2365	GAGATTTCGCTGAAGTTCAAGGACGATCAGGACCAATTTCTCTGATCTTCGCCCGCGCC	2306
QY	1312	TGGTTCAAGCTGACGACCGTGACCTTGGGCCCCAATACGTTTACATCGGCCCGGAAGTG	1371
Db	2305	TGGTTCAAGCTGACGATCGGCATATGGGCGGAAGTCCGTTTACGTCGGCCGGAATGT	2246
QY	1372	CCGGCAGAAGACCTGATTGGCAAGACCGGATTCGGCAGGTAACACCGACTACTCGCA	1431
Db	2245	CCGGCTGAAGACCTGATCGGCAAGATCCGATCCCGCAGGCTCCACGAGCTACGATGTC	2186
QY	1432	GAAGTGGTCAAGCAGAAAAATTGCACAAAGTGGCCTGAGCATTAAGTGAATGGTCTCCACC	1491
Db	2185	GCTGCGCTCAAGGCTAAGATCGCTGCTCGCGGCTTTCTGTCCCGCATCTGGTTTCAACC	2126
QY	1492	GCTTGGGACAGTCCCGTACTTATTCGGGTTCCGATATCGCGGGCGGTGCTAACGGTGC	1551
Db	2125	GCATGGGACAGTCCCGCACCTTTCGTTGGTTCGGACAAGCGCGCGCCCAATGGCGCG	2066
QY	1552	CGCATTCGCTTGGCCCCACAGAACGAGTGCAGGGCAACGAGCCGAGCGCCTTGGCGAA	1611
Db	2065	CGTATTCGCTCGACCGCAGAGGATTTGGGAAGGCAATGAGCGCCCGCTCTTTCCCGC	2006
QY	1612	GTGCTGAGGCTCTACGAGCAGATCTCTGCGCACACCGGGCTTAGCATCGCGGACGTGATC	1671

Db	2005	GTGCTTTCGGTCTGGAGCCGATGTCGCCGCAAAACGGTGCAGAGCATCGCCGATGTGATC	1946
Qy	1672	GTTCCTGCCCGTAGCGTAGCATCGAGAAAGCCGGAGAACACACAGGTTACGATGTGCGC	1731
Db	1945	GTTCCTGGCTGGCAATACGCGGTGGAGCGCGGCAAGCGCGCTGTTTCGATATCGCC	1886
Qy	1732	GTTCCTCTCTGAAAGCCGTGGCGATGCGACCGCGAGATGACCGACGAGACTCCTTC	1791
Db	1885	GTGCGCTTCGCGCGCGTCTGGTGTACGCTTCGCCGCGAGACAGCATCGGACAGCTTT	1826
Qy	1792	GCACCGCTGGAGCGCTGGCCGATGGCTTCGCAACTGGCAGAAAGAGATGTGGTG	1851
Db	1825	GCGCCGCTTCAGCGCGTGGCGGATGGTTTCGCAACTGGGTGAAGAAGGACTATGTCTC	1766
Qy	1852	AAGCGGAAGAGATGCTGCTGGATCGTGGCCACTGATGGGCTTAACCGGCCCGGAAATG	1911
Db	1765	AGCCCCGAAGAGCTGTGCTCGATCGGGCACAGCTTCTTGGCCTCACCGCGCGGAATC	1706
Qy	1912	ACGCTGCTGCTGGCGGTATGCGCGTACTGGGCACCAACTATGCTGGCAGCAACACGGC	1971
Db	1705	ACGCTGCTCATCGCGGCGCTTCGCGGTATCGCGGCCAATACGGCGGTGGCGGCATGGC	1646
Qy	1972	GTATTTCACCGATTGTGAAGGCCAGTGTGACCAACGACTTTTTTGTGAACCTGACCGCATG	2031
Db	1645	GTCTTCACCGATAAGCGCGGGCGCTTACAAACGACTTCTTCACGAGCTTGACGGACATG	1586
Qy	2032	GGNACAGCTGGAAGCGGTAGTAGCAAGCCCTACGAATTCGCGACGCGCAAGACCGGT	2091
Db	1585	GCCTATTTCCTGGGTCCCGACCGGCAACAATCTCTATGAGATCGGTGATGCGAAGACCGC	1526
Qy	2092	GCGTGAAGTGGACCGCTCGCGGTGGATCTGGTATTTGGTTTCAACTCGCTACGCGC	2151
Db	1525	GCAGCCAGATTTCGGCAACCGCGTGCATCTCGTATCGGCTTCAACTCCATCCTTCGCC	1466
Qy	2152	TCTTACGCAAGTGTACGCCAGGACGATAACCGCGAGAGTTCGTCAGAGACTTCGCTC	2211
Db	1465	GCTTATGCGCAAGTTATGCGCAGGACGACCAACAGGAGGAAAAATTCGCCCGGACTTCAT	1406
Qy	2212	GCGCGCTGGACCAAGTGATGAACCGCGACCGTTTCGA	2249
Db	1405	GCCGCTGGACGAAGGTGATGAACCGCGACCGTTTCGA	1368
RESULT 5			
AE009392			
LOCUS	AE009392	10299 bp	DNA linear BCT 20-DEC-2001
DEFINITION	Agrobacterium tumefaciens str. C58 linear chromosome, section 162 of 187 of the complete sequence.		
ACCESSION	AE009392 AE008689		
VERSION	AE009392.1 GI:17743133		
KEYWORDS			
SOURCE	Agrobacterium tumefaciens str. C58 (U. Washington)		
ORGANISM	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.		
REFERENCE	1 (bases 1 to 10299)		
AUTHORS	Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E., Chen,Y., Wood,L., Kitajima,J.P., Okura,V.K., Almeida Jr.,N.F., Zhou,Y., Bovee Sr.,D., Chapman,P., Klentzenning,J., Deatherage,G., Gillet,W., Grant,C., Guenther,D., Kutyavin,T., Levy,R., Li,M., McCrelland,E., Palmieri,A., Raymond,C., Rouse,G., Saenphimmachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I., Karp,P., Romero,P., Zhang,S., Too,H., Tao,Y., Biddle,P., Jung,M., Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C., Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V. and Nester,E.W.		
TITLE	The genome of the natural genetic engineer Agrobacterium tumefaciens C58		
JOURNAL	Science 294 (5550), 2317-2323 (2001)		
MEDLINE	21608550		
PUBMED	11743193		
REFERENCE	2 (bases 1 to 10299)		

AUTHORS

Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E.,
Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F.,
Zhou, Y., Bovee Sr., D., Chapman, P., Clendinning, J., Deatherage, G.,
Gillet, W., Grant, C., Guenther, D., Kutayavin, T., Levy, R., Li, M.,
McClelland, E., Palmieri, A., Raymond, C., Rouse, G.,
Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I.,
Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M.,
Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C.,
Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V.
and Nester, E.W.

Direct Submission

Submitted (27-SEP-2001) Department of Microbiology, University of
Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA
98195-7242, USA

FEATURES

source

Location/Qualifiers

1. .10299
/organism="Agrobacterium tumefaciens str. C58 (U.
Washington)"

/mol_type="genomic DNA"

/strain="C58"

/db_xref="taxon:180835"

121. .795

/gene="copR"

/note="synonym: Atu4638"

121. .795

/gene="copR"

/note="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer"

/codon_start=1

/transl_table=11

/product="two component response regulator"

/protein_id="AAL45432.1"

/db_xref="GI:17743134"

/translation="MRLLIEDERKEMADLSATLGKQIVIDHTMRIGDAMELTPROHL
VDALLDRRLPQEGGLAFIPKLRAGVDTPILLTANNEPEDRLAGLDGGADDYLGKP
FLVCELLARVRLRPSGLPAEIRAGRIDVDPHLHSLVINSIPFLDPRRELIVLA
LAKKGSVLSRLTEAAVYVEEIQSNALDAHLSRLKRLDASAGVSNIRIGIGY
LKKEDA"

792. .2126

/gene="copS"

/note="synonym: Atu4639"

792. .2126

/gene="copS"

/note="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer"

/codon_start=1

/transl_table=11

/product="two component sensor kinase"

/protein_id="AAL45433.1"

/db_xref="GI:17743135"

/translation="MKAQGSNPSLWQLSQISIVVSAMIAVIVGLCVYGLMLSP
NIALWGLSDALDESLSLDPKGLVVDGPMRLALTAENGTLMPAVSTLDGRVATYGA
VPTAYAEISRYLYLTADIRGGKGTAEGAVESLETWGPVRVWFGSGSTHVRQAFLT
LLAETKIYVPLAVILPAVFFSVPRIVROALAGLSVKKKAPILDPRRGSKLPVED
VPREVPLILSFNSILRSEQFQARQRRLIDAHLELTPAIQRIEGLPGPGRQR
LLADVRLGETAEGLAEPFRNDRRRVLDITRVVADLAPLALSAGYDIAFE
TAVVAYELQGNPFTLPRAINSVIRNAIDHGGKGMHIVSLADGEVVISDGGQIAED
QQQRIFFETPTPRSTGAGLSLVKQIVAIHGGRVLESSAFGSTFRLRF"

2162. .3109

/gene="Atu4640"

2162. .3109

/gene="Atu4640"

/note="putative; ORF located using Glimmer"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAL45434.1"

/db_xref="GI:17743136"

/translation="MLPQFWLSYRAGKHINATGVRTVQQWNNLLPFAFGRCRRSRG
AAAFATLAIIMSCSVPLKEGGTLTSAQFSPAKGFTKSRFTVDAYGLAGIKTVA
IVPTTFSPASSRVTSKDRILVSNALDRAICVALSDKYRTAALGQPADMTVTVVTD
LVPTNKWAGSVTAVTIGSGFVLPVPRLPVGLGLAVAEADVAGVQRAAAVWSK
GANSITKPRVSEIGDAIGLANDFSNTFARILLVKGVSEGLDLSVPSGHRIRSRALGGR

gene

CDS

PKYECDAYGRSGRGLQGVADKFGAPPETWTKKHARASGR"
complement(3171. .4079)
/gene="oxyR"
/note="synonym: Atu4641"
complement(3171. .4079)
/gene="oxyR"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, Lysr family"
/protein_id="AAL45435.1"
/db_xref="GI:17743137"

/translation="MIALSMKHLRYFDALAKIGHFGRAEACAIQSOPALSQIRELEE
LIGAPLVGRSQIRLTALGDEFAERTAILRSVDELDQDLARAGHGLPSGRIGTVIP
TVAPYLLQVTKTTRHYPGLEAPREAVTOKLLEDLRLDMLAIVALPVSEPALEE
VPLSEFEILVRPMEDAGMPVPSADKLGEMELLLELLEHCHCFRNGALSCTTNPAPRV
LMEGSSLTLYQVAGAGITVILFQMAVDMETRLSTVSVFRLAEPRFSRTIGIVWRKS
NPLSAQFAHISEIVRECGQLQKGLAP"

gene

CDS

/gene="kta"

/note="synonym: Atu4642"

4288. .6459

/gene="kta"

/note="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer"

/codon_start=1

/transl_table=11

/product="catalase"

/protein_id="AAL45436.1"

/db_xref="GI:17743138"

/translation="MDATSKPAGKCPVMHGNTASGKSVTEWPNALNLDLHQHDTK
TNPLGTFNTREALTKLDVEALKADRALMTDSQWMPADWSYGMARVTHAAGS
VRVTDGKGANTQREAPLNSWPNVNTDKRLLWPICKKYGNKISWADLIALAGT
IAYDVAGLKTGFGAFGRDIWAPKDYVWGEKWLAPSGCYGDVSKPTLENPLAA
VOMGLIYVNEPVGNGSDPLATAQMRTEFARMGDEETVALTAGGHTIGKSHGNS
AANLSPERAGPEIQGLVNTIKGRIGRTVYVSGEGAWTSEPTKWDNGFFDMLFK
HEWTLTHSPAGASQWAPITIAEDKPDVEDASIRTPMTDADMLKVDPIYREISL
KFDDQDHFSDVFAFWFKLTHRDMPKSRVYGPDPFAEDLIWQDPIPAGSTVDVAA
VKAKIAASGLSVADIVSTAMDSTARTFGSKRGKGANGARILWQDKWEGNEPARLSR
VLSVLEPIARETGASIADIVIAGNYVEQAAKAGFDIAPVFAAGRGDASAEQTDAD
SFAPLEPLADGFRNWKDYVSPSELLLDRAQLLGLTAPELTVLIGSLRGVYKNGG
AAHGVFDKPGALITDFTLTLDMAYSWVPTGNLIEIRDKRTGAARYSATRVLDLVIG
SNSILRAYAEVYAQDDNREKFAWTFIAAWTKVMNADREFDLI"

gene

CDS

/gene="Atu4643"

complement(6556. .6732)

/gene="Atu4643"

complement(6556. .6732)

/note="putative; ORF located using Glimmer"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAL45437.1"

/db_xref="GI:17743139"

/translation="MNEKKEAGDFADKKREQGRGVIVAGPDHENEKSGKTPAGPTE
KSSPERQVPSDDR"
complement(6855. .7724)

gene

CDS

/gene="Atu4644"

complement(6855. .7724)

/gene="Atu4644"

/note="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer"

/codon_start=1

/transl_table=11

/product="transcriptional regulator, AraC family"

/protein_id="AAL45438.1"

/db_xref="GI:17743140"

/translation="MAFQNHNAGQATLARIILNSLRVSEVKLSVNAHLITLAAAGH
AQVTDGEAFTLPAPGLAWLPAGHASRLRLSAGSRASLTSTTEIGLAHMTGTSAII
LRNLGROFTLELAEKDBELAGLYLETIANENLKPATETITNLSVTILRICOA
TTEIGAVSLPSGWTPEFVLLVAQHKCEHWSVDDYARDLGISDRILGSIKNATGLSPQ
AYIHRELLSEADLLNLSLQVAETAFRLGFDQPGFYFNRFRTGEGKSGSPGRRAAMR
AODIPLPSYAAW"

```

gene       7848..9299
/ gene="vdh"
/ note="synonym: Atu4645"
CDS        7848..9299
/ gene="vdh"
/ note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="vanillin: NAD oxidoreductase"
/protein_id="AAL45439.1"
/db_xref="GI:17743141"
/translation="MTVNLFIGGDHCAANKRFRNNPITGDVTSAAAAAIVEDAT
NLADAAAAPPKWSATTPGERRHLLAAALAAAGPQIIQAMMERIGATGAAGFNV
MAADMLVAAASLTHITGEVPSNRPFGTMALROPAGVYVSMAPWNPVILGVRSI
ATPLACGNTVVKMISLCPRTALIIIEAVASAGLPKGLVNAISNAPDDADIVETILIA
HPAVRNVTGSFVRGRVIAETAGRYLKPALLEGKAPFIVLDDADIDAAVAAAAG
AVYNQGLCMSTERRIIVLDSIADTFVEAPAKKASLTAGDPREGKPLGLSVSAEAS
RIRLVDDAVSKGARRVAGGGDGTMLDAIVDQVTPAMRLYSEESFPGVPVSIIRAGS
IDVLSIANESEFLGAATVFGDQARALSAARIESGICHVNGPTVHDEAQLPFGVK
ASGYRFGGTAGIAEFTELRWVTLQDPLHYPI"
complement(9326..10222)
/ gene="Atu4646"
complement(9326..10222)
/ gene="Atu4646"
/ note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"

Query Match      48.5%; Score 1098; DB 1; Length 10299;
Best Local Similarity 69.6%; Pred. No. 1.7e-219;
Matches 1488; Conservative 0; Mismatches 650; Indels 0; Gaps 0;

QY 112 GGTAAATGCTCTCATGACAGTGGTAAACACCTCGACCGGTACTTCCACAAAGATTGG 171
DB 4312 GGCAAGTGTCCCTCATGATGATGAGGCAATACGGCTCCGGCAATCGGTGACCGAATGG 4371

QY 172 TGGCCGGAAGGTTGAACCTGGATATTTTGCATCAGCAAGATCGAAATCAGACCCGATG 231
DB 4372 TGGCCGAACCGCTAAACCTCGACATCTGATCAGCAGCAGCAGCAACCAATCCGCTC 4431

QY 232 GATCCGGATTCAACTACCGTGAAGAAGTACGCAAGCTCGATTTCGACGGCTGAAGAAA 291
DB 4432 GGCACTCTTCAACTACCGGAAGCGCTGAACACGGTTGATGTCGAGCCCTCAAGGCC 4491

QY 292 GATGTCACCGGTTGATACCGATAGCCAAAGAGTGGTGGCCGCTGACTGGGGGCACTAC 351
DB 4492 GATCTGCGCGCTTATGACCGACAGCCAGGAATGGTGGCGCGCGACATGGGGCAGTTAT 4551

QY 352 GCGCGTTTGATGATCGGTATGGCTTGGCACTCCGCTGGCACTACCGTATTCGTATGGC 411
DB 4552 GTCCGGATGATGCCCGCTGTACTGGCATGGCCGCGGTTCCTATGCTGTACAGACGGT 4611

QY 412 CGTGGGGGGGTGGTACCGGAACCGAGCGTTTTCACCGCTCAACTCTCGGCCGGAAC 471
DB 4612 CGCGGGCGGCCCAATACCGGCAACAGGTTTTCACCGCTCAATTCCTGGCCCGGAAC 4671

QY 472 GTCAGCCTGGATAAGCGCGCTCTGCTGTGGCCGATCAAGAAGAGTACGCAACAAA 531
DB 4672 GTCAACACCGACAAAGGCGCGCGCTGCTGTGSCCGATCAAGAAGAAATACGCAACAG 4731

QY 532 ATCAGCTGGGACAGCTGATGATTCGTGGCTGGCAACCGTGGCTTATGAGTCCATGGCGTTA 591
DB 4732 ATTCTCTGGCCGACCTTATCGGCTCGCGGCGACGATCGCTTATGACGCTCGCGGCTG 4791

QY 592 CTTGCTTACGGCTTCTTTCGGCCGCGTGCATATTTGGGAACCCGAAAAGATATCTAC 651
DB 4792 AAGACCTTCGCTTCGCCCTCGCGCGGAAGACATCTGGCGCGGAAAAGGACCTAT 4851

QY 652 TGGGGTGACGAAAAGAGTGGCTGGCACTTCTGACGAACCGCTTACGACGCTGGAACAAG 711
DB 4852 TGGGGTGACGAAAAGAGTGGCTGGCGCGGAGCGCGGCTTATGCGGACGTGAGCAAG 4911

QY 712 CCAGACACCATGGAACCCGCTGGCGGCTGTGCCAATGGGTCTGATCTATGTGAACCCG 771

```

```

DB 4912 CCCGAGACGCTGGAAAACCCGCTTGGCGCGTGCAGATGGCGCTGATCTACGTCAACCCG 4971
QY 772 GAAGGTGTTAAACGCGCACCTGATCCGCTGAGAACCGCACAGAGGTACTTTGAACCTTC 831
DB 4972 GAAGGTGTTAAACGCGCAAGTCCGATCCGCTGGCGACGCGCGCGAGATGCGCGAACCCTT 5031
QY 832 GCCCGTATGCGGATGAACGACGAAACCCGACGCTTCACAGCTGGCGGCCACACCGTC 891
DB 5032 GCCCGATGGGATGGATGACGAGGAACCGTTGCCCTGACGGCGCGGCCACACCATC 5091
QY 892 GGTAAATGTCACGGTAAATGGCAATGCTCTGCGTTAGCCCTGACCCCAAAAGCCTCTGAC 951
DB 5092 GGCAAGTCCCATGGCAATGGCTGTGCCAATCTCAGCCCCCATCCGGAAGCTGTGGC 5151
QY 952 GTTGAACACAGGCTTAGGTGGGCAACCCCAACATGCAAGGAGGCAACCAACGCGC 1011
DB 5152 CCGGAATATCAGGGTCTCGGCTGATCAATACCAAGGCGCGCGCATTTGCCGTGACACC 5211
QY 1012 GTACCTCCGGTATCGAAGGTGCTTGGACACCAACCCACGAAATTCGATATGGGCTAT 1071
DB 5212 GTGGTGTCCGGTATCGAAGGCGCATGGACAAGCAACCAACCAAGTGGGACAACGGCTC 5271
QY 1072 TTCGACCTGCTGTTCCGCTACAAATTTGGAACTGAAAAGAGTCTCTGCCGTGCCACCAT 1131
DB 5272 TTCGACATGCTGTTCAAGCACGAGTGGACCTTCAGCACAGCCCGCGGTGATCGCAA 5331
QY 1132 TGGGAACCGATTGACATCAAAAAGGAAACCAAGCGGTTTACGCCAGCACCGCTCTATT 1191
DB 5332 TGGCGCGGATTTACCATCGCGGAAGAAGAACACCTGTTGATGTCGAGGATCGGTGAT 5391
QY 1192 CGCCACAACCCGATCATGACCGATGCGGATATGGCATAAAGTAAATCCGACCTATCGC 1251
DB 5392 CGCACCATCCGATGATGACCGACGCGACATGSCCTGAAGTTCGATCCGATCTACCGC 5451
QY 1252 GCTATCTGCAAAAATTCATGGCCGATCTCTGAGTACTTCAAGAAAATTTTCGGGAAGCG 1311
DB 5452 GAGATTTCGCTGAAGTTCAAGAGACGATCAGAGACATTTCTCTGATGTTCTTCGCGCGCGC 5511
QY 1312 TGGTTCAAGTACGACGACCGTACCTGGCGCCGAAATACGTTTACATCGGCCCGGAGTG 1371
DB 5512 TGGTTCAAGTACGACGATCGCGACATCGGGCCGAAAGTCCCGTTACGTCGGCCCGGATGT 5571
QY 1372 CCGGCAAGACCTGATTTGGCAAGACCCGATTTCCGGCAGGTAAACCCGACTACTCGCAA 1431
DB 5572 CCGGCTGAAGACCTGATCTGGCAGGATCCGATCCCGCAGGCTCCACGAGCTACGATGTC 5631
QY 1432 GAAGTGTCAAGCAGAAAATTCACAAAAGTGGCCTGAGCAATTAAGTGAATGGTCTCCACC 1491
DB 5632 GCTGCCGTCAAGGCTAAGATCGCTGCTCCGCGCTTCTGTCGCGCATCTGTTTCAACC 5691
QY 1492 GCTTGGGACAGTCCCGTACTTATCGCGTTTCGATATGGCGCGGTGCTTAACGGTGC 1551
DB 5692 GCATGGGACAGTCCCGCACCTTTCGCTGGTTCGGAACAGCGCGCGGCGCAATGGCGG 5751
QY 1552 CGCATTCGCTTGGCCCCACAGAGTGGCAGGCAACGAGCGGAGCGGAGCCCTGGGAAA 1611
DB 5752 CGTATTTCGCTCGCACCGCAGAGGATTTGGGAAGGCAATGAGCCCGCTCTTCCCGC 5811
QY 1612 GTCTGAGGCTTACAGCAGCAGATCTTGCAGCAACCGCGCTAGCATTCGCGGACGTGATC 1671
DB 5812 GTCTTTCGCTTTCGAGCGGATTTGCCCGCAACCGGTGCAAGCATTCGCGCATGATC 5871
QY 1672 GTTCTGGCGGTAGCTAGGCATCGAAGAACCGCGAAAGCAGCAGTACGATGTGCGC 1731
DB 5872 GTTCTGGCTGGCAATTTACGGCTGGAGACGCGCGGAAAGCGGCTGTGTTTCGATTCGCC 5931
QY 1732 GTTCCCTTCTTAAAGCCGCTGGCGATGGACCGCGAGATGACCGACGACGACTCTTTC 1791
DB 5932 GTGCCCTTTCGCGCGCGTCTGTGTGAGCGTTTCGCGCGAGCAGAGCATGCGGACGCTTT 5991
QY 1792 GCACCGCTGGAGCCGCTGGCCGATGGCTTCCGCAACTGCGCAGGAAGAGATGTGGTG 1851
DB 1851

```

Db 5992 GCGCGCTTGAGCGCTGGCGGATGGTTTCGCCAACTGGGTGAAGAGGACTATGTGCTC 6051

QY 1852 AAGCCGGAAGAGATGCTGCTGGATCGTGGCAGCTGATGGCTTAAACGGCGCCGAAATG 1911

Db 6052 AGCCCGAAGAGCTGCTGCTGATCGGCGACAGCTTCTTGGCCTCACCGCGCGGAATC 6111

QY 1912 ACCGTGCTGGCGGATGCTGGGCTACTGGGCAACCACTATGTTGGCAACCAACAGCGG 1971

Db 6112 ACCGTGCTGATCGCGGCTGCTGGGCTCATCGGCGCCCAATTAACGGCGGTGGCGCGCATGGC 6171

QY 1972 GTATTACCGGATTTGAAGCGCAGTTGACCAACAGCACTTTTGTGTAACCTCAGCGGATATG 2031

Db 6172 GTCTTACCGATAGCGGGGCGCTTACAGGCACTTCTTACAGAGCTTACAGGATGACGAGATG 6231

QY 2032 GGGAAACAGCTGGAAGCGGTTAGGTAGCAACGCCCTACGAAATCCGCGACCGCAAGACCGGT 2091

Db 6232 GCCTATTCTGGTCCGACCGCGCAACAATCTCTATGAGATCCGTGATCGCAAGACCGCG 6291

QY 2092 GCCGTGAGTGGACCGCTCGCGGTGATCTGATTTGGTATTTGTTTCAACTCCTACTGGCG 2151

Db 6292 GCAGCCAGATATTCCGCAACCGCGTGGATCTCGTATCGGCTCCAATCTCATCTCTGGC 6351

QY 2152 TCTTACGAGAGTGTACGCCAGGAGATTAACCGCGAGAGTTCGTTCAGAGACTTCGTC 2211

Db 6352 GCTTATGCGGAGTTTATGCGAGGACGACAGAGGNAATTTGCGCCGCGACTTCATT 6411

QY 2212 GCCGCTGGACCAAGTATGATGAACCGCGACCGCTTTTCCA 2249

Db 6412 GCCGCTGGAGAGGTGATGAACCGCGACCGCTTTTCCA 6449

RESULT 6

AB033631 2640 bp DNA linear BCT 09-FEB-2000

LOCUS Agrobacterium tumefaciens kta gene for catalase, complete cds.

DEFINITION Agrobacterium tumefaciens

ACCESSION AB033631

VERSION AB033631.1 GI:6691547

KEYWORDS catalase.

SOURCE Agrobacterium tumefaciens (Rhizobium radiobacter)

ORGANISM Agrobacterium tumefaciens

REFERENCE Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium. 1 (sites)

AUTHORS Xu,X.Q. and Pan,S.Q.

TITLE An Agrobacterium catalase is a virulence factor involved in tumorigenesis

JOURNAL Mol. Microbiol. 35 (2), 407-414 (2000)

MEDLINE 20118000

PUBMED 10652101

REFERENCE 2 (bases 1 to 2640)

AUTHORS Xu,X.Q. and Pan,S.Q.

TITLE Direct Submission

JOURNAL Submitted (18-OCT-1999) Xiuqin Xu, National University of Singapore, Biological Sciences, 10 Science Drive 4, Singapore 117543, Singapore (E-mail:scip7311@nus.edu.sg, Tel:658743596)

COMMENT Sequence updated (22-OCT-1999).

FEATURES

source

1..2640

/organism="Agrobacterium tumefaciens"

/mol_type="genomic DNA"

/db_xref="taxon:358"

308..2479

/gene="kta"

308..2479

/gene="kta"

/function="hydrogen peroxide detoxification"

/codon_start=1

/transl_table=11

/product="catalase"

/protein_id="BA089349.1"

/db_xref="GI:6691548"

/translation="MDATSKPAGKCPVHHGGNTVSGKSVTEWPNALNLDLHHDYTK

TNPLGTSFNREALKTLIDVEALKADLRALMTDSQEWNPADMGSTVGMWARTVHAAGS

YRVTDGGGANTGNRFAPLNSWPNVNTDKRELLWPIKKYKNKISWADLIAGT

IAYDAGLKTGFAGREDIWAPEKDTYWGDEKWLAPSDGRYGVSKPELEPNLAA

VOMGLIYVNPGEVNGKSDPLATAAOMRETFARMGMDDEETVALTAGGHTICKSHNGS

AANLSPDPEAAGPEYGLGWINTKRGIGRDTVVSGLEGAMTSEPTWKGDFDMLFK

HEWLTSPAGASQWAPITIAEDKGPVDVEDASIRTPMMDADMAUKVADYTRISL

AFKQDHFSDVFAFMFKLTHRDMGPKSRVYGPVPAEDLIWDQDPIPAQSTSYDVA

VKAKIAGSLGSLVADLVADMSARTFRGSRGGANGARILAPKQWEGNEPARLSR

VLSVLEPTARETGASIADIVLAGNYGVEQAAGAAGFIDIAFFAAGDASAGTADAD

SPAPLEPLADGFRNWKDYVVSPELLLDRAQLGLTAPBLTVLIGLRYIGANYGG

AAGVFTDKPGALTDTDFTTLTDMAYSWVPTNNLYETRDRKTKGAARYSATRVDLVIG

SNSTILRAVEVYAODDNRKFAFDIFIAWTVMNADRPDLI*

BASE COUNT 547 a 790 c 800 g 503 t

ORIGIN

Query Match 48.5%; Score 1096.4; DB 1; Length 2640;

Best Local Similarity 69.6%; Pred. No. 4.1e-219;

Matches 1487; Conservative 0; Mismatches 651; Indels 0; Gaps 0;

QY 112 GGTAAATGCTCTGTCATGCACGGTGGTTAAACCTCGACCGGTACTTCCAAAGAAATGG 171

Db 332 GCGAAGTGTCCCGTCATGCATGGAGCAATACGGTCTCCGCAAAATCGGTACCCGAATGG 391

QY 172 TGGCCGGAAGGTTGAACCTGATATTTTGCATCAGCAAGATCGCAANTCAGACCCGATG 231

Db 392 TGGCCGGAAGGCTAAACCTCGACATCTGCATCAGCACAGACACCAATCGCGTC 451

QY 232 GATCCGATTTTCAACTACCGTGAAGAGTACGCAAGCTCGATTTCCAGCGCTGAAGAA 291

Db 452 GGCACCTCTTCAACTACCGGGAAGCGCTGAAGCGCTTGATGTGAAGCCCTCANGGCC 511

QY 292 GATGTCCAGCGTGTGATGACCGATAGCAAGAGTGGTGGCCCGCTGACTGGGGGACATPAC 351

Db 512 GATCTGCGCGCTTATGACCGACAGCCAGGAATGGTGGCCGCGGACTGGGGCAGTTAT 571

QY 352 GCGGTTTGTGATCGGTATGGCTTGGCAGCTCCGCTGGCAGCTACCTACCTGTTGCTGATGGC 411

Db 572 GTCGCGATGTGGCCCGGTGTACCTGGCATGCGCCCGGTCTCTATGTGTGTCAGACCGT 631

QY 412 CGTGGGCGGTGTACCGGAAGCGCAGCGCTTTGACCGCTCAACTCTCTGCGCGGACAAAC 471

Db 632 CGCGCGCGCCCAATACCGGCAACAGCGTTTGGACCGCTCAATCTCTGCGCGGACAAAC 691

QY 472 GTCAGCTGTGATTAAGCGCGCTGCTGTGTGGCGGATCAAGAAGTACGGCAACAAA 531

Db 692 GTCACACCGACAGAGCGCGCGCTGCTGTGGCGGATCAAGAAGAAATACGGCAACAAG 751

QY 532 ATCAGCTGGCAGACCTGATCTGTGCTGCGACCGTGGCTTATGCTCCATGGCTGCTG 591

Db 752 ATTCTGGCGGACCTTATCGGCTCGCGCGGACGATCGGCTATGAGCTGCGGGGTCTG 811

QY 592 CTTGCTTACGGCTTCTCTTTCCGCGCGCTCGATATTTGGGAACCCGAAAAAGATATCTAC 651

Db 812 AAGACCTTTCGCTTCGCTTCGCGCGCAAGACATCTGCGCGCGGAAAAAGACACCTAT 871

QY 652 TGGGGTGACGAAAAAGAGTGGCTGCGACCTTCTGTAGCAACGCTAGCGGAGCTGAACAAAG 711

Db 872 TGGGGTGACGAAAAAGAAATGGCTGGCGCGGACGCGCGCTTATGGCGAGCTGAGCAAG 931

QY 712 CCAGAGACCATGGAAAAACCCGCTGCGCGCTGCTCCAAATGGCTGATCTATGTGAACCCG 771

Db 932 CCCGAGACGCTGGAAAAACCCGCTTGGCGCGCTGAGATGGGCGCTGATCTACGTCAACCCG 991

QY 772 GAAGTGTTAAACGGCCACCTGTATCCGCTGAGAACCGCACAGCAGCTACTTGTGAACCTTC 831

Db 992 GAAGTGTTCAACGGCAAGTCCGATCCGCTGGGAGCGGCGCGCAGATCGCGGAACCTTT 1051

QY 832 GCCCGTATGGCGATGAACGACGAAAAAACCCGACGCCCTTACAGCTGGCGGCGCACACCGCTC 891

Db 1052 GCCCGCATGGGATGGATGACGAGGAAACCGCTTGGCTTACGCGCGCGGCGCACACCATC 1111

QY 892 GGTAAATGCTCAGGTAAATGGCAATGCTCTGCTGGTTAGCCCTGACCCCAAAAGACCTCTGAC 951

Db 1112 GCGAAGTCCCATGGCAATGGCAGTGTCTCCCAATCTTCAGCCCGCGATCCGGAAGCTGCTGCGC 1171

QY	952	GTGAAACACGAGCGTTAGTGTGGGCAACCCCAACATGCAGGCGCAAGGCAAGCAACGCC	1011
DB			
DB	1172	CCGGAATATCAGGCTCTCGGCTGGATCAATACCAAGGCGCGGCACTTGGCGGTGACACC	1231
QY	1012	GTGACCTCGGTATCGAAGGTGCTTGACCAACCAACCCAGCAAAATTCGATATGGGTAT	1071
DB			
DB	1232	GTGTGTGGGTATTCGAAGGCGCATGCAAGCAACCAAGCAACCAAGTGGGACAAGCGTTC	1291
QY	1072	TTGACCTGCTGTTCGGCTACAATTTGGAACTGAAAAGAGTCTCGCGGTGCCACCAT	1131
DB			
DB	1292	TTGCACATGCTTTCAAGCAGGAGTGACCCCTGACGCAACGCCCGCGGTGCAATCCGAA	1351
QY	1132	TGGAAACCGATTTGACATCAAAAGAGGAACCAAGCGGTTGACGCCAGCAGCCCTCTATT	1191
DB			
DB	1352	TGGCGCGGATTACCATCGCGCAAGAACAGACAGGCTGTGATGTCGAGGATGGGTGATC	1411
QY	1192	GCCACAAACCGGATCATGACCGATGCGGATATGCGGATAAAGGTAATTCGACCTATCGC	1251
DB			
DB	1412	GCCACCATTCGGGATGATGACCGAGCGCGACATGCGCCTGAAGTGTGATCTATACCG	1471
QY	1252	GCTATCTGCGAAAAATTCATGGCGGATCTCTGAGTACTTCAAGAAAACTTTCGGGAAGCG	1311
DB			
DB	1472	GAGATTGCGTGAAGTTCAAGAGAGATCAGAGCCATTCTCTGATGCTTTCGCCCGCGCC	1531
QY	1312	TGTTCAAGCTGACGCCCGGTGATCGGCGCCGAAATCAGGTTACATCGGCGCGGAAGTG	1371
DB			
DB	1532	TGTTCAAGCTGACGCATCGGACATGGCGGAAGTCCGTTACGTCGGCGCGGATGTT	1591
QY	1372	CCGCAAGACCTGATTTGGCAAGACCCCGATTCGCGCAGGTAAACCCGACTACTGCGAA	1431
DB			
DB	1592	CCGCTGAAGACCTGATCTGGAGATTCGATCCCGCAGCGCTCCAGCAGCTACGATGTC	1651
QY	1432	GAGTGTCTCAAGCAGAAATTTGCACAAGTGGCTGAGCATTAAGTAGAGATGTTCTCCACC	1491
DB			
DB	1652	GCTCCCTCAAGCTAGATCGCTGCGCCTTCGGCTTCTGTCGCCGATCTGTTTCAACC	1711
QY	1492	GCTTTGGACAGTGGCCCTACTTATTCGCGGTTCCGATATGCGCGCGGTGCTAAACGTTGCC	1551
DB			
DB	1712	GCATGGACAGTGGCCGACCTTCGCTGGTTTCGGACAAGCGCGCGCCCAATGGCGG	1771
QY	1552	CGCATTCGTTGGCCCCACAGAAGCTGCGAGGGCAACAGCGCGAGCGCTTGGCGAAA	1611
DB			
DB	1772	CGTATTGCTCTCGCACCGCGAAGGATTTGGGAAGGCAATGAGCGCGCTCTTTCGCCG	1831
QY	1612	GTGCTGAGCGCTACGAGCAGATCTCTGCGCACACCGCGCTAGCATCGCGGAGTGATC	1671
DB			
DB	1832	GTGCTTTGCTGTTGGAGCGATTCGCCCGCAACCGGTGCAAGCATTCGCCGATGTGATC	1891
QY	1672	GTTCGCGCGGTAGCGTAGGCTATCGAAGAACCGCGAAGCAGCAGGTTACGATGTGCGC	1731
DB			
DB	1892	GTTCGTGCTGCAATTAACGCGTGAGCAGCGCGCAAGCGGCTGTTTCGATATCGCC	1951
QY	1732	GTTCCTCTCTGAAGCGCGTGGCGATGCGACCGCGCGAGATGACCGACGAGACTCTCTC	1791
DB			
DB	1952	GTGCGCTTCGCGCGCGTCTGTGTGAGCTTCTCGCGCAGCAGCGATGCGCAGCGTTT	2011
QY	1792	GCACGCTGGAGCGCGTGGCGATGGCTTCGCAACTGGCAGAGAAAGATGATGGTG	1851
DB			
DB	2012	GCGCGCTTGAGCGCGTGGCGATGGTTTCGCAACTGGGTGAAGAGGACTATGTGCTC	2071
QY	1852	AAGCCGGAAGAGATGCTGCTGGATGCTGCGCAGCTGATGGGCTTAACCGCGCGGAAATG	1911
DB			
DB	2072	AGCCCCGAAGAGTCTGCTCGATCGCGCACAGCTTCTTGGCCTCACCGCGCGGAACCTC	2131
QY	1912	ACCGTGTGCTGGCGGTATGCGGTACTGGGCAACCAACTATGGTGGCACAACACAGCGC	1971
DB			
DB	2132	ACCGTCTCATCGCGCGCTCCGCTCATCGGCGCCAAATACGCGCGTGGCGGCATGGC	2191
QY	1972	GTATTACCGATTGTGAAGCCAGTGTGACCAACGACTTTTTTGTGACCTGACCGATATG	2031
DB			
DB	2192	GTCTTACCAGTAAGCCGGGGCGCTTACAAACGAGCTTCTTACAGAGTTGACGGACATG	2251

QY	2032	GGAAACAGCTGGAAAGCGGTAGTAGCAACCGCTACGAAATCCGCGACCGCAAGACCGGT	2091
DB			
DB	2252	GCCTATTCTCTGGGTCCGACCGCAACATCTCTATGAGATCGTGTATGCAAGACCGGC	2311
QY	2092	GCCTGAAGTGGACCGCTCGCGGTGGATCTGGTATTTGGTTCGAACCTCGCTACTGCGC	2151
DB			
DB	2312	GCAGCCAGATATTTCGGCAACCCCGCTCGATCTCTGATCGGCTCCAACCTCCATCTCTG	2371
QY	2152	TCCTAGCAGAACTGACGCCAGGACGATAACCGCGAGAAAGTTCTGTCAGAGACTTTCGTC	2211
DB			
DB	2372	GCCTATGCGGAAGTTATGCGCAGGACGACACACGAGGAAAATTCGCCGCGACTTCATT	2431
QY	2212	GCCGCTCGACCAAGTGTATGAACCGCGACCGCTTTCGA	2249
DB			
DB	2432	GCCGCTGGACGAAGTGTATGAACCGCGACCGCTTTCGA	2469
RESULT 7			
LOCUS	AP005086	296100 bp	DNA linear BCT 05-MAR-2003
DEFINITION	Vibrio parahaemolyticus DNA, chromosome 2, complete sequence, 3/6.		
ACCESSION	AP005086	BA000032	
VERSION	AP005086.1	GI:28608836	
KEYWORDS	Vibrio parahaemolyticus		
SOURCE	Vibrio parahaemolyticus		
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Vibrio.		
REFERENCE	1		
AUTHORS	Nasu, H., Iida, T., Sugahara, T., Yamaichi, Y., Park, K. S., Yokoyama, K., Makino, K., Shinagawa, H. and Honda, T.		
TITLE	A filamentous phage associated with recent pandemic Vibrio parahaemolyticus O3:K6 strains		
JOURNAL	J. Clin. Microbiol. 38 (6), 2156-2161 (2000)		
MEDLINE	20295086		
PUBMED	10834969		
REFERENCE	2		
AUTHORS	Makino, K., Oshima, K., Kurokawa, K., Kurokawa, K., Yokoyama, K., Uda, T., Tagomori, K., Iijima, Y., Najima, M., Nakano, M., Yamashita, A., Kubota, Y., Kimura, S., Yasunaga, T., Honda, T., Shinagawa, H., Hattori, M. and Iida, T.		
TITLE	Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism distinct from that of V. cholerae		
JOURNAL	Lancet 361 (9359), 743-749 (2003)		
MEDLINE	22508454		
PUBMED	12620739		
REFERENCE	3 (bases 1 to 296100)		
AUTHORS	Oshima, K., Kurokawa, K., Makino, K., Yokoyama, K., Yasunaga, T., Honda, T., Shinagawa, H., Hattori, M. and Iida, T.		
TITLE	Direct Submission		
JOURNAL	Submitted (09-APR-2002) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail: kenken-gen-info.osaka-u.ac.jp),		
COMMENT	URL: http://www.gen-info.osaka-u.ac.jp/, Fax: 81-6-6879-2047 genome project This clone was isolated from a patient presenting with acute gastroenteritis.		
FEATURES	Location/Qualifiers		
source	1. .296100		
	/organism="Vibrio parahaemolyticus"		
	/mol_type="genomic DNA"		
	/strain="O3:K6"		
	/sub_strain="RIMD 2210633"		
	/db_xref="taxon:670"		
	/chromosome="2"		
gene	57. .1031		
	/gene="VPA0602"		
CDS	57. .1031		
	/gene="VPA0602"		
	/note="Similar to GB:AAA25445.1 (J04233) percent identity 27 in 299 aa"		
	/codon_start=1		


```

/evidence-not_experimental
/transl_table=11
/product="putative transcription activator Nahr"
/protein_id="BAC61945.1"
/db_xref="gi:28808837"
/translation="MLRESLDMIENYLLKPLALLLETKSITATANOMSTSASVSRT
LRSREIFQDELLTRKNGRMELTPKALISIRAKLANIITHDELFEDMSFSDSDIKI
NIANNSTIAYWLADPIIYFSENAPNIELTIDWNELTDPDRINHLVHFGHYPLML
KSNLVQRKVKDMVYCRKDPHINNESISETIASYPLAIHLQHWNERQDKLSRL
LNSHDIPKLVKTKTHINVIINAVETTDLLPFCSIYIAEQLEKEFSYITSSDEVLPQLQ
KEKFGFYDKAYKNDPMVWLHSTISKLMDFLDVYNNKFDSDREKSR"
complement(1056..1146)
/gene="trna-Ser-6"
complement(1056..1146)
/gene="trna-Ser-6"
/product="trna-Ser"
/note="anticodon: TGA"
1476..2156
/gene="VPA0603"
1476..2156
/gene="VPA0603"
/note="similar to GB:AAF96706.1 (AE004409) percent
identity 72 in 236 aa"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="NodN-related protein"
/protein_id="BAC61946.1"
/db_xref="gi:28808838"
/translation="MKVTNLFKDKVDALSAHQSEFMNWSPTLROYSDFLGKAQNSF
LFSNVRBHQPNVNEEVAQPAPVQLVPEAQKLDLSGKIGEVHFGDWVHVVSQERI
NQGFADQEDMWITDERASVESPFKTIHAGFELTALLPLKLDSDVDADKPLPPTAK
LVNVLGNQVRFPYVKGAGNNVRAVSTLTKVTPIKKGLEIERIKVEIEGVRRPGAVV
VSVTLQHF"
complement(2210..3037)
/gene="VPA0604"
complement(2210..3037)
/gene="VPA0604"
/note="similar to GB:AAF96705.1 (AE004409) percent
identity 61 in 272 aa"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="ABC transporter, periplasmic substrate-binding
protein"
/protein_id="BAC61947.1"
/db_xref="gi:28808839"
/translation="MLRFTVATLLAMAVSLPSAKAEVNIISGSTSVARVMDVLAEEYN
KTHPDNVIYVQIGSSAGIHWNGVYKLGSSRYLSESEKEDLNVPPIAYDGLAVV
VNRNVSNTLSAQQLFDIYKGEIRKNKEVGADQPIAVVTREASGSRYPSESILGLT
KIINDRLVSDISPNLNVNSMYKTIYNHNTAIGTISVGSVDLSVKAIQFNIEPT
AANTANKYKLARPLVLYKVDLSLQAGKDFVAFKSEAGQKTIADYGYIPVKNFNQ"
3249..3647
/gene="VPA0605"
3249..3647
/gene="VPA0605"
/note="similar to GB:AAF96704.1 (AE004408) percent
identity 37 in 108 aa"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC61948.1"
/db_xref="gi:28808840"
/translation="WALGCIKTTYQENVMSVSOMATSRLYKLVQILCSARNANLPVE
QGKLYEPIENGVELHHAHFLDQSHSYTSPIAKILFDSOTQLWRFTVPASRSEDIR
WIPYKKLPHSHTLDELAELESDFQACFE"
complement(3652..4455)
/gene="VPA0606"
complement(3652..4455)
/gene="VPA0606"
/note="similar to GB:AAG16667.1 (AF255769) percent
identity 37 in 110 aa"

```

```

/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="putative AraC-type regulatory protein"
/protein_id="BAC61949.1"
/db_xref="gi:28808841"
/translation="MPNIEIIRFNHFTARKSERVYATNGHLYVVEEGALVHVHNGEQ
FELAQAGTLYNSDLRSAAEIPGENGFKALVFDISLFCFEKKAHPGLHSEAEHRR
FYFSPSNTSITQLKNTLLALARNAPDYQSHIAMALLSLMVEQDILSIIDDA
SLTASQAIKVIENKIEKIDITLEGIAEHMSIATLKRLLAAENLSFSQILKVRINY
AATOLRVQSKSITIEAFESGFKSAHFSTAFKSYINITPKDFRQVVRG"
complement(4527..6533)
/gene="VPA0607"
complement(4527..6533)
/gene="VPA0607"
/note="similar to GB:AAF96703.1 (AE004408) percent
identity 81 in 655 aa"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="exoribonuclease II"
/protein_id="BAC61950.1"
/db_xref="gi:28808842"
/translation="MFQDNPLLAQLKQIQENLPPKKESIRATDKGFGFLEVDKTSF
FIPPAYMKCKIHGDKVVAIIRTEREVAEPQELIEQLTRFIRGVKRFKGLNVVPD
HPWKLKSLKAKLKGKLPDNFAEGDWYVAHLVRHPLKGDNTFFVEISEKITDADKI
APWWTLAQNDLPNSEPAGIENWELKDDADLERIEMTHVPFTIDGESTKMDDALYA
KKTESGDFELTIAJADPTAYITPEDEMDKVARERGYTYLPGRNIPMLPDLADNLCS
LIEGEIPAICCTVTVSKDVGIGDDIKFAANIKSHARLAYDHWISDWLENSDAWQP
SEETATVRDIYEFESLARAEWRKNVVPDPDTRFELSDNDVIAIHADMRKSANR
LVESMITANICAGTIREKPEFGYVFNTHAGLKPEKIEEVQLVNPEGTLETFEASIA
TLEGFAALRRWLAVQETSYLDNRIRKFOAYSEVGNQPLPHYAMGLDIYATVTSIRKY
GDMINHRMLKAVILDKEPVQDDQVEELALHRKHKKIAERNYSDWLYARTLADEPS
KOTCTGTGIFDINRAGARVLLGAAAFIPGALILDNKEKIECNGDNGTISIDKEVV
YKLGDTLEIVLADVNVQENRSLVAKPTQVFADQAPQTEQTYSEE"
6913..8844
/gene="VPA0608"
6913..8844
/gene="VPA0608"
/note="similar to GB:AAF96702.1 (AE004408) percent
identity 79 in 656 aa"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="ATP-dependent RNA helicase Dead"
/protein_id="BAC61951.1"
/db_xref="gi:28808843"
/translation="MQDSVIOFSDLSLNDLSILSDGMGFVSPPIQAAAIPLHLEGA
DALGKAQGTGKTAAFSLPLNKLKLDQKQPOAIVLAPTRELAIOVAAEMKNLGNKIK
GLKVLIEYGGASIVDOMRALKNGARVVVGTGQVDLLINRERLHLDVETVFLDEADE
MLNMGFVDDVTAIMEHAPESARQVLFSAIEMPMKLNIVERLDRDPTVDVAGKNHTVD
KVQQQFWVVKGEKDEAMSRLEETEASIVFVTRDTRDELADMLSARGFKAALH
GDIPQSLEERTVDHIKQGVIDILVATDVVARGLDVPRTHVYNDIPDVSRYIHRIG
RTGRAGRKGAILLVHTQIRMLRTIERVTKSSMEETQLPLRDQVAERLNLKLAEL
AEKHEKALDKFAELVEKIQTSLEIDPMAAALLKROQKGRPLFYIGSDPVAEIERD
KORRERREGGERSNNQDMDTQIQVREGQVQVQKIVGALANLGLGKSGIGA
IKLAQGHFTVOLPKAMTSEAAKSLKIRIQQDYGAVVCDDFDPSRGGRDGRD
GRRGEGFGRNGRNGRREGGREGGFRGNGRNGRNGRNGRNGRNGRNGRNGRNGR
RGNVGERGCHGRRTQDA"
complement(8919..10520)
/gene="VPA0609"
complement(8919..10520)
/gene="VPA0609"
/note="similar to GB:AAK90145.1 (AE008359) percent
identity 36 in 431 aa"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC61952.1"
/db_xref="gi:28808844"
/translation="MPTIAFLVILGLILLCWGMQMTNCKTSKTEVPQGTWLLMLISC
FITAYAAFLHAIKMPKVASLLETGVSULLFGGSIFVLVIKWSNDSILELHVASBREK

```

Query Match	48.0%	Score 1085;	DB 1;	Length 296100;	
Best Local Similarity	69.2%;	Pred. No. 7.5e-217;			
Matches 1481;	Conservative 0;	Mismatches 660;	Indels 0;	Gaps 0;	
QY	112	GGTAAATGCTCTGTCATGCACGGTGGTAAACACCTCGACCGGTACTTCCAAACAAGATGCG 171	DB	1132	TGGGAACCGATTGACATCAAAAAGGAAAAACGCCGTTGACGCCACAGCCCTCTATT 1191
DB	194643	GGTAAATGTCAGTATGTCATGTTGGTGAACAACATCTACAGACAAGTCGGTGATGGATTTGG 194702	DB	195663	TGGGAGCCTATCGACATTGAAGAACAAGACAGCGCTGTTGATGTAGAAGATGGCTCTAAG 195722
QY	172	TGGCGGAAGGGTTGAACCTGGGATATTTTGCATCAGCAAGATCGCAAAATCAGACCCGATG 231	QY	1192	CGCCACACCCGATCATGACCGGATGCGGATATGGCGATAAAGTAAATCCGACCTATCGC 1251
DB	194703	TGGCCGAATGCCTTAAACTTAGACATCTGCATCAGCAGCACTCCAAAGCAATCCATTTT 194762	DB	195723	CGCTTAAATCAANTATGACCGATGCGCATATGSCACTCAAAATGGACCCCTGAATATCGC 195782
QY	232	GATCCGGATTCAACTACCGTGAAGAAGTACGCAAGCTCGATTTTCAGCGCGCTGGAAGAAA 291	QY	1252	GCTATCTCGCAAAAATTCATGGCCGATCCTGAGTACTTCAAGAAAATCTTTCGCGAAGGCG 1311
DB	194763	GGTCCAGATTTTAAATTAAGAAGAACTTTAAAAAACTTGACCTCGACGCACTCAAGCAA 194822	DB	195783	AAGATTTTCAGAGCGCTTCCACACAGACCTTGCACTACTTTTGAAGATATGTTTGGCCGTGG 195842
QY	292	GATGTCACGCGTTGATGACCGATAGCAAGAGTGGTGGCCGCTGACTGGGGGCACTAC 351	QY	1312	TGGTTCAGCTGAGCGCACCGTGACCTGGGCCCGGAAATACAGTTTACATCGGCCCGGGAAGTG 1371
DB	194823	GACTTAAAGACCTAATGACAACAGCCCAAGTTTGGTGGCCTCGACATGGGACACTAC 194882	DB	195843	TGGTTCAAACTGACTACCGCTGATATGGACCAAAATCGTGCTACCTTAGSCCCAGATGTA 195902
QY	352	GGCGGTTTGATGATCCGTATGGCTTGGCACTCCGCTGGCACCTACCCTATTGCTGATGGC 411	QY	1372	CCGCGAGAAGACCTGATTTGGCAAGACCCGATTTCCGCGCAGGTAAACCGCACTACTCGGAA 1431
DB	194883	GGCGGTTTGATGATTCGTATGGCATGSCACCGCGCTGGTAGCTATTCGTATCGCAGACGGA 194942	DB	195903	CCTAGCAAGATCTGATTTTGGCAAGACCCACCGCAGCAGGCAAAAACGGACTACAACGTC 195963
QY	412	CGTGGGGCGGTGTTACCGGAAGCCAGCGCTTTTGACACCGCTCAACTCTTGGCCGGACAAC 471	QY	1432	GAACTGCTCAAGCAGAAAAATTCACAAAGTGGCTGAGCATTTAGTGAGATGGTCTCCACC 1491
DB	194943	CGTGGTGGTCCGCAACCGGTAAACAAGCTTTGGCCCTCTCAACTCTTGGCCTGATAAC 195002	DB	195963	GATCTTGTAAGAGGCAAAATCGAAGCGAGCGCTTGCTATCTGCTGATTTGGTGGCCACC 196022
QY	472	GTCAGCCTGGATAAAGCGCGCGCTCTGCTGTGCCGATCAAGAAGAAAGTACGCAACAAA 531	QY	1492	CTTTGGGACAGTGCCTACTTATCGCGGTTTCGATATGCGCGCGCGGTGCTTAACGGTGGC 1551
DB	195003	GCCAACTAGACAAGGACGCGCACTCTTTTGGCCGATCAAGAGAAGTACGGCAACAAA 195062	DB	196023	CTTTGGGACAGTGCCTACTTATCGTGCTCTGACCGAGCTGGCGCGCAAAATGGCGGG 196082
QY	532	ATCAGCTGGGAGACCTGATGATCTGGCTTGGCCACCGTGGCTTATGAGTCCATGGGCTTA 591	QY	1552	CGCATTCGCTTGGCCCCACACAGACGAGTGGCAGGGCAACGAGCGGAGCGCTTGGCGAAA 1611
DB	195063	TTAAGCTGGGCGACCTGATATTTCTCGCAGGTAAACATGGCTTATGAATCGATGGCTTT 195122	DB	196083	AGAATTCGTTTGGCTCCGCAAAAAGATTTGCCAAGGTAAACGAGCCAGAGGCTTAAAGTCGT 196142
QY	592	CCTGCTTACGCTTCTCTTTTGGCGCGCTCGATATTTGGGAACCCGCAAAAAGATCTAC 651	QY	1612	GTGTGACGCGTCTACGAGCAGATCTCTGCCGACCCGCGCTAGCATCGCGGACGTGATC 1671
DB	195123	AAAACCTTTGGCTTTGGTTCGCGAAGATATTTGGCATCCAGAAAAGACACTAT 195182	DB	196143	GTTCTTGCTGTGTAGAAAAGCATCGCAGCAGAAAGGGCTGTAGCGTGTGCTGATGCTATC 196202
QY	652	TGGGGTGACAAAAGAGTGGCTGGCACCTTCTGACGAACCGGTACGCGCAGCTGGAACAAG 711	QY	1672	GTTCTGGCGCGGTAGCGTTAGGCATCGAGAAAACCGCGCAAGCAGCAGGTACGATGTGCGC 1731
DB	195183	TGGGGTTCGAGCAAGAGTGGCTTGGCCAAAGTGGCGGAAACACAGCCGTTACTCCGGA 195242	DB	196203	GTACTTGTGTGTAACGTAGGTATTTGAATTTGGCTTGGCGAGCGGCGACGACCATGTATCT 196262
QY	712	CCAGAGACCATGGAACCCGCTGGCGGCTGTCCAAATAGTGGTCTGATCTATGTGAACCCG 771	QY	1732	GTTCCCTTCTGTAAAGCCGCTGGCGATGCGACCGCGGAGATACCGACGACAGACTCTCTTC 1791
DB	195243	GAGCGTGATTTAGAGAACCCACTTGGCAGCAGTATGATGGGCTTGATCTATGTGAACCCA 195302	DB	196263	GTGCTTTTCGCACACCGTCCAGCGGATGCTTCTCAAGATATGACGATGTGGNAATGCTTC 196322
QY	772	GAAGTGCTTAACGGCCACCTGATCCGCTTGAGAACCGCACAGCAGGTACTTCAAAACCTTC 831	QY	1792	GCACCGCTGAGCGCGCTGGCCGATGGCTTCCGCAACTGGCAGAAAGAGATATGTGGTG 1851
DB	195303	GAAGGGTGATGGCAACCCAGATCCATTTGAANAACGGCCGCAAGACATGCGAGTGACATTT 195362	DB	196323	GAAGTGCTTAAACCTGTGGCTGACGGTTTCCGAAATTTGGTTGAAGAAGACTATGTGGTG 196382
QY	832	GCOCGTATGCGGATGAACAGCAAAAAACCGCAGCCCTCACAGCTGGCGGCCACACCGCTC 891	QY	1852	AAGCGGAAGAGATGCTGTGGATCGTGGCACCTGATGGCTTAACCGCGCCGGAATG 1911
DB	195363	GCGCGTATGCAATGAACGATAGGAACCGTTTGCACCTAACGGCTGGTGCATACCGGTG 195422	DB	196383	ANGCCGAAGAGTTGTTGCTCGACCGAGCGCACCTATGGGTTTGACAGCACCAAGAAATG 196442
QY	892	GGTAATTTGTCACGGTAATGCAATGGCTCTGGGTTAGCCCCCTGACCCCAAAAGCCTTGAC 951	QY	1912	ACCGTGCTGTGGCGGATATCCGCTACTGGGCACCAACTATGGTGGCACCAACACAGCGC 1971
DB	195423	GGTAAATGTCACGGTAACGGTGATGTGCCAACGCTTGGCCCCAGATCCAGAAAGCGCGCTGAC 195482	DB	196443	ACGCTACTGATTTGGTGGTTTACGCGTCTGGTGTAGCAACTACGCTGTAGCAAAAGATGCG 196502
QY	952	GTTGAAAACACAGGCTTAGTGTGGGCAACCCCAACATGCAGGGCAAGGCAAGCAACGCC 1011	QY	1972	GTATTCACCAATGTGAAGCCGAGTTGACCAACGACTTTTTTGTGAACTGACCGATATG 2031
DB	195483	GTGCATGAACAAGGCTTGGTTGGATGAACCAACAGACTCTGGTATTTGGTGTGACAGG 195542	DB	196503	GTCTTTTACCCACCGCTGGGTAGCTTATCGAAGATTTCTTCGTAACCTCACCGATATG 196562
QY	1012	GTGACCTCGGGTATCGAAGTGCTTGGACACACAAACCCACAGAAATTCGATATGGGCTAT 1071	QY	2032	GGGAACAGCTTGAAGCCGCTAGGTAGCAACGCTACGAAATCCGCGACCGCAAGACCGGT 2091
DB	195543	GTAAACAGCGCTTTGGAAGGGCTTGGCAACGCAATCCGCACTCAATTTGGGATTAACGTTAT 195602	DB	196563	GCATTACATATGGAAGCCAGCGGTTGAGAACCAATACGAAATTCGTGATCTGAACAGAC 196622
QY	1072	TTGACGCTGCTTTCGCTACAAATGGGAAGTGAAGAGTCCCTGGCGGCTGCCACCAT 1131	QY	2092	GCCGTGAAGTGGACCGCCTCGCGGGTGGATCTGGTATTTTGGTTCCAACTCGCTACTCGGC 2151
DB	195603	TTCTATTGCTCTTCAAAATACGATTTGGGAACCTGAAGAAGGCCACGCGCGCTTGGCAA 195662	DB	196623	CAAGTGAATGGACCGCGACTCGAGTGGATCTTGTCTTTGGTTTCAAACTCGATTTCTACGT 196682
			QY	2152	TCATTACCAAGTGTACGCCAGGACGATTAACGGCGAGAGTTGCTGTAGAGACTTCGTC 2211
			DB	196683	TCGTATCCGGAAGTGTATGCGCAAGATGATAATCAGGAAAAATTCATCATGATTTCCGTC 196742
			QY	2212	GCCGCTTGGACCAAAAGTGAACCGCGACCGTTTTCGAGGT 2252

Db	196743	GCTGCATGGACGAAGTCAATGAAACGGCGATTCGTGTTTGACTT	196783
RESULT 8			
AF197161			
LOCUS	2700 bp	DNA	linear BCT 14-SEP-2000
DEFINITION	Synechococcus PCC6301 catalase-peroxidase gene, complete cds.		
ACCESSION	AF197161		
VERSION	AF197161.1	GI:6180192	
KEYWORDS			
SOURCE	Synechococcus sp. PCC 6301		
ORGANISM	Synechococcus sp. PCC 6301		
REFERENCE	Bacteria; Cyanobacteria; Chroococcales; Synechococcus.		
AUTHORS	Engler,M., Regelsberger,G., Jakopitsch,C., Furtmuller,P.G., Rucker,F., Peschek,G.A. and Obinger,C.		
TITLE	Nucleotide sequence analysis, overexpression in Escherichia coli and kinetic characterization of Anacystis nidulans catalase-peroxidase		
JOURNAL	Biochimie 82 (3), 211-219 (2000)		
MEDLINE	20322938		
PUBMED	10863004		
REFERENCE	2 (bases 1 to 2700)		
AUTHORS	Engler,M., Regelsberger,G., Jakopitsch C., Furtmueller,P.G., Rucker,F., Peschek,G.A. and Obinger,C.		
TITLE	Direct Submission		
JOURNAL	Submitted (21-OCT-1999) Institute of Chemistry, University of Agricultural Sciences, Muthgasse 18, Vienna A-1190, Austria		
FEATURES	location/Qualifiers		
source	1..2700		
	/organism="Synechococcus sp. PCC 6301"		
	/mol_type="genomic DNA"		
	/strain="PCC 6301"		
	/db_xref="taxon:1139"		
	/note="synonym: Anacystis nidulans"		
CDS	241..2403		
	/EC_number="1.11.1.6"		
	/note="hydroperoxidase"		
	/codon_start=1		
	/transl_table=11		
	/product="catalase-peroxidase"		
	/protein_id="AAF05841.1"		
	/db_xref="GI:6180193"		
	/translation="MTATRGKCPVMEHGATTNVAEWPWKALNDILSQHDKTNP MGDPFNYQEVKKLDYLAALKDQLQALMDSODWMPADWGHYGLMLRLTWHAAGTVRI ADKGAGGTGNQRFAPLNSWDNTLDKARLLWPIKOXYGNKLSWADLIAYATVI ESKGLTFEGAFREDIWIHPKDIYWGPKEWVPPTNSRVTGDBRELENPLAATVM GLIYNPEGVDNPDLKTAHDVRVFARMNADEETVALTAGGHTVGKCHNGNAL LKPEGEADVDEGLIMINKTQSGIRNAVTSLEGAVTPHTPOWDNGYFRMLLNTDW ELKSPAGAWQSEFP INFEEDLPIDVEDPS IRRNLVMTDADMAMKDPEYRKISERFY QPAYFAVDFARFWFLTRDMGPKARY IGPDVPQEDLIWQDIPAGNRINDYOAVKD RIAASGLSISELYSTAWSARTYNSDKRGGANGARLRAPOKRDWEGNEPDRLAKVLA VLSEIARAATCASADVIVLAGNVGVQQAARAAGEIVLPPAPGRGDATAEQTDESFA VLEPIHDGYRNWLKQDPAPEELLORTQLLGTEPMFTVLGGILRVLTGNHGGTGH GVSTFDREGLVITDNFFNLDTMDNLTLMKPAGKNLYEICDRKTNOVKWTATRVDLVFGFSNS ILRAYSELYAQDDNKKEKFVERFDAWTKVMNADRFDLD"		
BASE COUNT	702 a	627 c	754 g
ORIGIN			
Query Match	47.6%	Score 1076;	DB 1; Length 2700;
Best Local Similarity	68.7%;	Pred. No. 7.6e-215;	
Matches 1481;	Conservative	0; Mismatches 675;	Indels 0; Gaps 0;
Qy	99	GAGAGCTATGACGGGTAATGTCCTGTCATGCACGGTGTGAACACCTCGACC GG TACTTC	158
Db	243	GACAGCAACCGGGGTAATGTCGGGTCAATGCACGGCGGAGCAACACCGCTTAATATTC	302
Qy	159	CAACAAGAATGTGGCGCGGAAGGGTTGAACCTGGATATTTTGCATCAGCAAGATCGCAA	218
Db	303	GACTGCTGAGTGGTGGCCAAAGGCACTCAACCTGGATATTTTGAGCCGACCATCGCAA	362
Qy	219	ATCAGACCCCGATGGATCCGGATTTTCAACTACCGTGAAGAAGTACGCAAGCTCGATTCGA	278

Db	1443	GTTCGACGGCTTGGTTCAAGTTAACCCACCGCGGATATGGGCGCGAAAGCCCTTACAT	1502
Qy	1359	CGCCCGGAAGTGGCGCAGAACACCTGATTTGGCAAGACCCGATTCGGCGAGGTAAACAC	1418
Db	1503	TGGCCCGGATGGCCACAGGAAGACCTGATTTGGCAGATCCATTCGGCGGCGCAACCG	1562
Qy	1419	CGACTACTGGAGAAGTGGTCAAGCAGAGAAATTCACAAGTGGCTGAGCATATAGTGA	1478
Db	1563	CAACTATGACGTGCAAGCGGTGAAGATCGGATTCGTCGCCAGTGGCATAGTATCACTGA	1622
Qy	1479	GATGGCTCCACCGCTTGGGACAGTGGCGGTCTACTTATCGGGTTCCTATCGCGGCGG	1538
Db	1623	GCTAGTCAGCACAGCTTGGGATAGCGCCGCTACTTATCGAAATTCGATAGCGGGCGG	1682
Qy	1539	TGCTAACCGTCCCGCATTCGCTTGGCCCGCACAGACGAGTGGCAGGCAACGACGCGGA	1598
Db	1683	GGCGAATGGCGACGATTCGGTTAGCACCCAGAAAGATTTGGAAAGGATGAACCCGA	1742
Qy	1599	CGCGCTGGCAAGTGTAGCGCTCTACGAGCAGATCTCTGCCGACACCGCGGCTAGCAT	1658
Db	1743	TCGCTTGGCAAGTGTGCTGTCTGGAAGGAATTCGCTCCGCTACTGGGCGGAGCGT	1802
Qy	1659	CGCGAGCTGATCTTCTGCGCGTAGCGTAGGCATCGAGAAACCGCGCAACGACGAG	1718
Db	1803	TGCGAGCTCATTTGCTGTGCTGCAATGTTGGGTGGAGCAGCAGCAGGCTGCCGG	1862
Qy	1719	TTACGATGTGCGGCTTCCCTTCCCTGAAAGCCGCTGGCGATCGCAGCCCGCAGATGACCGA	1778
Db	1863	TGTTGAAATGTGCTTCCCTTTCGCGCGGCTGCTGGCGATGCAACGCTGACCAACCGA	1922
Qy	1779	CGCAGACTCTTTCGACCGCTGAGCGCTGGCGGATGGCTTCGCGCAACTGGCAGAGAA	1838
Db	1923	TACGGAATCTCTTGCAGTGTGGAGCGGATTCAGATGGCTATCGCAACTGGCTCAAGCA	1982
Qy	1839	AGATGATGTGTGAAGCGGAAGAGATGCTGCTGATCGTGGCGAGCTGATGGCTTAAC	1898
Db	1983	GGACTATGGCGCAACGCTGAAGATTCCTGTTGATCGCAGCGCAACTGTTGGGTCTGAC	2042
Qy	1899	CGCGCCGGAATACCGCTGCTGCTGGCGGATGTCGGGTACTGGGCAACCAACTATGCTGG	1958
Db	2043	GGCTCCAGAGATGACGCTGTTGATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2102
Qy	1959	CACCAACACGGGTATTCACCGATTTGGAAGCCAGTTGACCAACACACTTTTGTGAA	2018
Db	2103	TAGGAACACGGGTGCTTTCACCGATCCGGAAGGGGTGTTAACGAATGACTTTTTCGTGA	2162
Qy	2019	CCTGACCGATATGGGAACAGCTGGAAGCGGTAGGTAGCAACGCTTACGAATCCCGGA	2078
Db	2163	TCGACCGCATGATTAATTCCTGGAACCGGCTGGAAACAACTGATGAATCTCGGA	2222
Qy	2079	CCGCAAGACGGTCCGCTGAAGTGGACCGCTCGCGGGTGGATCTGGTATTTGGTCCAA	2138
Db	2223	TCGCAAGCAATCAGTGAAGTGGAGCGCAACGCGAGTCGATTTGCTTTGGATCAA	2282
Qy	2139	CTGCTACTCGCTCTTACCCAGAAAGTGTACGCCAGGAGCATACGCCGAGAGTTCGT	2198
Db	2283	TTGATTTCTCGAGACCTTACGAGCTCTATGCACAGACGACAAACAGAGAGTGTGT	2342
Qy	2199	CAGAGACTTCTGCGCCCTGGACCAAGTGAACCGCGGCTTTCACGCTCG	2254
Db	2343	CGGAGACTTGTGCTGCTCCCTGGAGGAAGTGAATGCCGATCGCTTGTATCTGG	2398
RESULT	9		
LOCUS	AE016806	304490 bp	DNA linear BCT 28-FEB-2003
DEFINITION	Vibrio vulnificus CMCP6 chromosome I section 10 of 11 of the complete sequence.		
ACCESSION	AE016806		
VERSION	AE016806.1		
KEYWORDS	GI:27362155		
SOURCE	Vibrio vulnificus CMCP6		
ORGANISM	Vibrio vulnificus CMCP6		

Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Vibrio.
1 (bases 1 to 304490)
Rhee,J.H., Kim,S.Y., Chung,S.S., Kim,J.J., Moon,Y.H., Jeong,H. and Choy,H.E.
Complete genome sequence of Vibrio vulnificus CMCP6
Unpublished
Jeong,H., Moon,Y.H. and Kim,J.J.
Direct Submission
Submitted (13-DEC-2002) Genotech Corp., 461-6, Jeonmin-dong, Yuseong-gu, Daejeon 305-811, South Korea
2 (bases 1 to 304490)
Rhee,J.H., Kim,S.Y., Chung,S.S., Lee,S.E. and Choy,H.E.
Direct Submission
Submitted (13-DEC-2002) Department of Microbiology, Genome Research Center for Enteropathogenic Bacteria, Chonnam National University Medical School, Hak-I-dong, Dong-Gu, Kwang-Ju 501-746, South Korea
Location/Qualifiers
1. 304490
/organism="Vibrio vulnificus CMCP6"
/mol_type="genomic DNA"
/strain="CMCP6"
/db_xref="taxon:216895"
/chromosome="I"
109..1602
/locus_tag="VV12664"
109..1602
/locus_tag="VV12664"
/note="contains a CheY-like receiver, AAA-type ATPase, and DNA-binding domains; COG2204"
/codon_start=1
/transl_table=11
/product="Response regulator"
/protein_id="AA011011.1"
/db_xref="GI:27362156"
/translation="MRKVLIVEDSTSLAIYKQYKDEPDYIFHVTGRTAQIFTER
NPQLVLIDKLPMDSGEDVLDINQNDIPTAVIATAHGSDVIAVSLMKNKGAEDFLE
KTKADRLKTSIAVHLKRAKLEHLVENMQSRFDRPRVHGFIGACLPQAVVYKTDVA
PTWAVFNGESGTGKEVCAEAHKQSLRKEQPIALNCAIPRDLMESIFGHLKGA
FTCATWRKGAAMLANGGTLFLDELCEMELEMKKLLRFLQTGTFTPLGSKVKVDV
RIICATNDRPLKEVEGRFREDLYIRVHVPIEMPPPLRERGNIDVELANHLFTKYAK
DGKFSISKEAQSLIKKNWPNVROLQNIIRNVLDNDQNLNHLVPAQLTTPQ
TVPEAKLSTPPQAPQAIHEMHNGHEALNHSLETQTSKANPLTHNAFHHSQGSIRP
MWQIERAIGNAIAFCGDNVISAAMVLELSPSTVYRKKQAEAEANA"
1595..2728
/locus_tag="VV12665"
1595..2728
/locus_tag="VV12665"
/note="COG0438"
/codon_start=1
/transl_table=11
/product="Glycosyltransferase"
/protein_id="AA011012.1"
/db_xref="GI:27362157"
/translation="MPDDNKATLKKRPAKSHRVGKREIMLLDSLVYGGIESHWIELA
SALKIAGHYVRVLLTKYVRRPIMQOLRQIRVNVISOLSGNHPMPIMQALVALKR
YRPAIVHTGYKANIYARLARMSFPHRLVSTYHAGESKTKLWLYDWDRJLTSEIS
DHRFAYSDAILHSLPAGERINNFVPVSELAKSHQNSQHIGFVGRISYKKAADRFVE
LADHFFSLFHYLDGPREPLETLANDNVHVEFGVSDMTKVMKPSIVLLIPSRFEG
PMAALEAMANGAVIAMNVGQLGQLEHGQNGIAQTADLADLHLSWLEMPSEAOOR
LLFQARETIMQHSYPOAVLPQILLSYHIDNAFPNLRKLTPHH"
2836..4029
/locus_tag="VV12666"
2836..4029
/locus_tag="VV12666"
/note="COG0438"
/codon_start=1
/transl_table=11
/product="Glycosyltransferase"
/protein_id="AA011013.1"
/db_xref="GI:27362158"
/translation="MTTRTPILFVHYGDNWIRGSETTLIHLATLDTQOPEPIWNSNC

Db 109998 GCGCGTGGTGGCGGATACCGGTAACCAACAGCTTTTCTCGCTCCGTTGAACCTCTTGGCCCTGAT 110057
 Qy 469 AACCTCAGCCTGGATAAGCGCGCGCTCTGCTGTGGCCGATCAAGAAGAAGTAGCGCAAC 528
 Db 110058 AACGCCAACCTTGATNAAGCGCGAGCTACTTTGGCCCAATTAAAGCAAAAGTAGCGCAAC 110117
 Qy 529 AAATACAGCTGGGACAGACCTGATGATTTCTGGCTGGCCACCGTGGCTTATGAGTCCATGGGC 588
 Db 110118 AAAATAGCTGGGCTGATTTGATGATCTTGGCGGTAACTATGATGGCTACGAATCAATGGGT 110177
 Qy 589 TTACCTGCTTACGCTTCTCTTTCGCGCGCTGATATTTTGGGACCCGAAAAGATATC 648
 Db 110178 CTGAACACCTTTGGTTTTGCTTCGGGTGGAAGACATCTGGCATCCAGAAAGACATT 110237
 Qy 649 TACTGGGGTGACCAAAAAGAGTGGCTGGCACCTTCTGACGAACGCTACGGCGAGCTGAAC 708
 Db 110238 TACTGGGGTGACAGCAAGAATGGTTAGCGCCATCTGGTCAGNAACACGCGCTACTCT 110297
 Qy 709 AAGCCAGAGACCATGGAACCCCGCTGGCGGTGTCCAAATGGTCTGATCTATGTGAAC 768
 Db 110298 GGTGAACGCGACCTTAGAAAACCCGCTGCTGCTGATGATGGGCTTCATTTAGCTCAAC 110357
 Qy 769 CCGGAAGTGTAAACGCGCCACCTGTATCCGCTGAGAACCGCACAGCAGGTACTTGAAC 828
 Db 110358 CTTGAAGGTGTAGATGTATCCGACCCACTAAAACACGCAAGACATGCGGTGACT 110417
 Qy 829 TTCGCCGTATGGCGATGAACGAGCAAAAACCGCACCTCTACAGCTGGCGGCCACAC 888
 Db 110418 TTTGCTGCGATGGGATGAACGATGAAGAAACGGTTCGCTGACGGCGGGTGTACACG 110477
 Qy 889 GTGCGTAATTTGTCACGCTAATGGATGCTCTGCTGCTGTACCCCTGACCCAAAGCCTCT 948
 Db 110478 GTTGGTAAACACACCGTAAACGCGACGCAACCTTAGTGTGCGGATCCAGAAAGTGGC 110537
 Qy 949 GACCTTCAAAACACGAGCTTAGTGTGGGCAACCCCAACATGACGAGGCAAGCAAC 1008
 Db 110538 GATTGGAAGAGCAAGCGCTAGTGTGAACACCACTAAGTCGCGCGCATTTGTCGCGAC 110597
 Qy 1009 GCGGTGACCTCGGTATCGAAGTGTGTTGACCAACCAACCCCAAGAAATCGATATGGC 1058
 Db 110598 ACGGTGACCACTGATTTGAAGCGCATGGACGACAAACCAACACATGGGATAACGCG 110657
 Qy 1069 TATTTGACCTGCTGTTGCGCTACAAATTTGGGAAGTGAAGAGTCTCGCGGTGCCAC 1128
 Db 110658 TTTCTTCATCTGATTTAGCTACGATTTGTTGTTGATAGAGAGCCGACGAGTGTGG 110717
 Qy 1129 CATTTGGAACCGATTGACATCAAAAGGAAACCAAGCGGTGTGACGCCAGCCCTCT 1188
 Db 110718 CAATGGGAACCGATGAACATCAAGAGGAAGACAAACCGTTGATGGAAGATCCAACG 110777
 Qy 1189 ATTGCGCACACCGCATGATGACCGATGCGGATATGGCGATAAAGTAAATCCGACCTAT 1248
 Db 110778 ATTGCGCACACCCCATCATGACGATGACAGATGCGCTGAACTGGACCCAGAGTAC 110837
 Qy 1249 CGCGCTATCTGCGAATAATTCATGCGCGATCTGAGTACTTCAAGAAAATTTGCGGAAG 1308
 Db 110838 CGTAAGATTCTGAGCGATTCCAAAGATCCCGCTTACTTCTCTGAAACCTTCGCTCGC 110897
 Qy 1309 CGGTGTTCAAGTGTGAGCGACCGTGAACCTGGCGCCGAAATCACGTTACATCGGCCGGA 1368
 Db 110898 GCCTGGTTTAACTGATCAACCGGACATGGGACCAAAAGCTCGCTACTTTGGCCAGAT 110957
 Qy 1369 GTGCGCGCAGAAGACCTGATTTGGCAAGACCCGATTCGCGAGGTAAACCGCACTACTGC 1428
 Db 110958 GTCCCTGCAGAAAGTTGATTTGGCAAGATCCAGTTCGAGTGTGCTGAAGACTACGAC 111017
 Qy 1429 GAAGAAGTGTCAAGCAAGAAATTTGCAAAAGTGGCTGAGCATTTAGTGAAGTGTCTCC 1488
 Db 111018 GTAGATGCAATGAAGCGCAAAATCATCGCCAGTGGTTTGAAGCATCGCGCAATGGTTCA 111077
 Qy 1489 ACGCTGGGACAGTGGCCGCTACTTATTCGCGGTTCCGATATGCGCGCGGTGCTAAGCGT 1548
 Db 111078 ACCGCATGGGATAGCGCACGCACTTTCCGCAACTCAGATAAACCGCGTGTGCCACCGT 111137

Qy 1549 GCGCGCATTCGCTTGGCCACACAGAGTGGCAGGCAACAGCGCGGAGCGCTCGC 1608
 Db 111138 GCTCGTATTGACTTGGCCACAGAGATTGGTTAGCAACGACCAAGAAAATCGCG 111197
 Qy 1609 AAGTGTCTGAGGCTACAGAGCAGATCTGCGCAGACACCGGCGTAGCATCGGGAGCGT 1668
 Db 111198 AAGTGTCTTAAATGTCTTGAAGCAATTTGCTCTGATTCACATCAGTGTTCAGACACC 111257
 Qy 1669 ATGCTTCTGCGCGTAGCGTAGGCTAGGCAAGGACCGGAAAGCAGCAGGTAGCATGTG 1728
 Db 111258 ATGCTGTTCAGGTAATGTGGGTGTGGAAGACCGCGAAAGCGCTGTGTTGCAATC 111317
 Qy 1729 CGGTTCCTTCCTGAAAGCGCGTGGCGATGCGCAACCGCGAGATGACCGACGAGACTCC 1788
 Db 111318 ACGGTTCCTGTTGCAGCAGGTGCGGGGATGCGCACCATTTGACNAACCGGATGTGGAATCT 111377
 Qy 1789 TTGCGACCGCTGGAGCGCTGGCGATGCGTTCGCAACTTCGCGAAGAAAGATGTG 1848
 Db 111378 TTGCGATCTCTTGGCGGATTTGCAGATGCGTTCGCAACTTCGCGAAGAAACACTATGCG 111437
 Qy 1849 GTGAAGCGGAAGAGATGCTGCGATGCGTGGCAGCTGATGGCTTAAACGCGCCGAA 1908
 Db 111438 GTCACCTGAAAGATTTGCTGATGCGAGCTCAATTTAGTTTAAAGCGCCGAG 111497
 Qy 1909 ATGACCTGCTGCTGGCGGTATGCGGTACTGCGCACCACTATGTTGCGCAACCAAC 1968
 Db 111498 ATGACGCTATTGTTGTTGCTGTTGCTAGGTACCAACCATGCGGTACTATGACAC 111557
 Qy 1969 GCGGTATTACCGATTGTGAAGCGCAGTGTGACCAACGACTTTTTTGTGAACCTGACCGAT 2028
 Db 111558 GGTGTGTTTACCACAAATGTCGTGCACTAAGTAATGACTTCTTTGTGAACCTGACAGAC 111617
 Qy 2029 ATGGGGAACAGCTGGAAGCGGTAGGTAGCAACGCGCTACCAAAATCCGCGACCAAGACC 2088
 Db 111618 ATGCGTTACACCTGGAAGCGCAACCGGTGCGCAATTTTACGACATCTGCGAGCGTAACAGC 111677
 Qy 2089 GGTGCGCTGAAGTGGACCGCTCGCGGTGGATCTGTTATTTGTTTCCAACTCGCTACTG 2148
 Db 111678 GGCAATGTGAATGGACAGCTACTCGTGTGTTGACTTAGTGTGTTGTTTCCAACTCTATCTTG 111737
 Qy 2149 CGCTCTTACGACAGATGTAGCGCCAGGACGATAACGGCGAGAAAGTTTCGTCAGAGCTTC 2208
 Db 111738 CGTGTCTTACCTGAAAGTGTACGCTCAAGATGACAAAGAAATTCGTGGAAGACTTT 111797
 Qy 2209 GTGCGCCCTGGACCAAGTGTAGCGCGACGCTTTTCGACGTCGCGTCTGTA 2261
 Db 111798 GTGCGCGCTGGACTAAAGTAAGTAAGCAACCGGATCGCTTCCATCTAATCGAA 111850

RESULT 10 AP005085/c

LOCUS AP005085 303450 bp DNA linear BCT 05-MAR-2003
 DEFINITION Vibrio parahaemolyticus DNA, chromosome 2, complete sequence, 2/6.
 ACCESSION AP005085 BA000032
 VERSION AP005085.1 GI:28808465
 KEYWORDS

SOURCE Vibrio parahaemolyticus

ORGANISM Vibrio parahaemolyticus

Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;

Vibrionaceae; Vibrio.

REFERENCE

1

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

2

AUTHORS

1

Nasu, H., Iida, T., Sugahara, T., Yamachi, Y., Park, K.S., Yokoyama, K.,

Makino, K., Shinagawa, H. and Honda, T.

A filamentous phage associated with recent pandemic Vibrio

parahaemolyticus O3:K6 strains

J. Clin. Microbiol. 38 (6), 2156-2161 (2000)

20295086

10834969

2

AUTHORS

1

Makino, K., Oshima, K., Kurokawa, K., Yokoyama, K., Uda, T.,

Tagomori, K., Iijima, Y., Nakajima, M., Nakano, M., Shinagawa, A.,

Kubota, Y., Kimura, S., Yasunaga, T., Honda, T., Shinagawa, H.,

Hattori, M. and Iida, T.
 Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism
 distinct from that of *V. cholerae*
 Lancet 361 (9359), 743-749 (2003)
 MEDLINE 22508454
 PUBMED 12620739
 REFERENCE 3 (bases 1 to 303450)
 AUTHORS Oshima, K., Kurokawa, K., Makino, K., Yokoyama, K., Yasunaga, T.,
 Honda, T., Shinagawa, H., Hattori, M. and Iida, T.
 Direct Submission
 TITLE Information Research Center: 3-1, Yamadaoka, Suita, Osaka 565-0871,
 JOURNAL Japan (E-mail: ken@gen-info.osaka-u.ac.jp,
 URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365,
 Fax: 81-6-6879-2047)
 COMMENT genome project
 This clone was isolated from a patient presenting with acute
 gastroenteritis.
 FEATURES
 source Location/Qualifiers
 1..303450
 /organism="Vibrio parahaemolyticus"
 /mol_type="genomic DNA"
 /strain="O3:K6"
 /sub_strain="RIMD 2210633"
 /db_xref="taxon:670"
 /chromosome="2"
 81..1574
 /gene="VPA0300"
 81..1574
 /gene="VPA0300"
 /note="similar to EMB:CAC46187.1 (ALS91787) percent
 identity 68 in 487 aa"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=11
 /product="hypothetical signal peptide protein"
 /protein_id="BAC61643.1"
 /db_xref="GI:28808466"
 /translation="MNLKAVPKALLVTLTLLASSLPSSVAYAAPNANSSLVQQWQEGN
 WPTQNEAKLVQELYYQRAISAYTQMOPAMVIGNRDGSSEKFGKYNLVPTWDRMD
 SRTLVPTNADVYVSMNYLDLKEKPLVVAAPKPVIGMFTDFORTITDVLGIDPRA
 RGLVLLPPNVDGPPKGYFTFTSTNNVLFPTVMKKGGENGADPSDAVKIAROTR
 IYPLDEKNIQPKFPNASCORINMAYPTDFSWEKMKQVDEEIPSAISPELGVIL
 ASIGLVKGPNTVEEKQLLEKAVKAPKMTAQRLRKDERNLITDKROYERAWA
 GATSEYMOESLIDIDRAAFYQAYSSAPAMVMTMNAQSKYPTFTRDSKGDILNGSH
 QYKHLPANPAKLFMAVTLNLYVDGTPETSQLPSKNGFDNIEKNTDGSIDLYFGP
 QLPNGAPESNTIKTIPGRDFTAIRLYGTGIEFFDQTKWPDVLKIK"
 complement(1835..2314)
 /gene="VPA0301"
 complement(1835..2314)
 /gene="VPA0301"
 /note="similar to DBJ:BAF43758.1 (AP003138) percent
 identity 57 in 158 aa"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=11
 /product="putative peptide methionine sulfoxide reductase"
 /protein_id="BAC61644.1"
 /db_xref="GI:28808467"
 /translation="MMKKIYFAGGCLWGQEFMRHLPGVISTEAGRANGKTDNTQSEY
 DGYACVTFEDPSQVSIETLMGYFEIIDPYSINKQGGDDGKRYTGVYSDRYKHLE
 IAKSPIATRYDADKIAVEVLPLKNTVPSDEEHQDLTRFPNDYCHIPDLILLHYKNK"
 2493..2882
 /gene="VPA0302"
 2493..2882
 /gene="VPA0302"
 /note="similar to GB:AAE64344.1 (L16865) percent identity
 52 in 129 aa"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAC61645.1"

/db_xref="GI:28808468"
 /translation="MKIEHVAIWTERLEELKGFYEKYFNAYSNNDKYHNPKKHFSYFL
 SFEGARLELMSEGVTTSCNSHMQVTGLAHFAFGALGSEQAVDQITKTLVDDGYQRI
 DGPRTYGDYVESCVDPDGNGRIELTV"
 2974..3714
 /gene="VPA0303"
 2974..3714
 /gene="VPA0303"
 /note="similar to GB:AAL20622.1 (AE008775) percent
 identity 54 in 243 aa"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=11
 /product="putative regulatory protein"
 /protein_id="BAC61646.1"
 /db_xref="GI:28808469"
 /translation="MNPORNEILQIVNDRKRVQVTELSDIIGVSGVTIROLNLFLEQ
 GYLKRVHGAATALQSDDDIDTREVRFIDKQFLANKAADLVAPEVNETVLLEGSSANALLA
 RTLAERGDTVITITTSAYIAHLIRNTSANIILGGVYHQGESLVGLPKLCIENIHFS
 TAFIGIDGHDQDTGTSRDMRADIAEAILAKKRRNIVLTDSSKFGQIYPSIGKTNE
 ISVLLTDKAPKSDLEQLKLGVEVVLG"
 complement(3733..4089)
 /gene="VPA0304"
 complement(3733..4089)
 /gene="VPA0304"
 /note="similar to GB:AAF96321.1 (AE004376) percent
 identity 76 in 117 aa"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="BAC61647.1"
 /db_xref="GI:28808470"
 /translation="MHOEKLNYVEFGTPNIGATKAFEFHVGWQFVDYGPDYAAFSG
 QGLDGGFYSAEQVSTQNGAALLFYSSDIHATLEKVKAFGGQIIRPIFERPGGGRFH
 FTERSGNEFAVWSESH"
 4319..5860
 /gene="VPA0305"
 4319..5860
 /gene="VPA0305"
 /note="similar to REF:NP_231225.1 (NC_002505) percent
 identity 83 in 510 aa"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=11
 /product="catalase"
 /protein_id="BAC61648.1"
 /db_xref="GI:28808471"
 /translation="MVSTHWRATMQMSKSLITVGLASTLSLQAOTLFRDNGAPVGDN
 QNSITAGENGSVLLQDVHLIQLFARERIPERYVHARGTGAHGEFVASGDFSLTV
 SAPFTKGVTPVVFVRFSTVIHSKGSPELTPRDPGRFATKFTTECGNMDLVGNLPPVF
 IRDSIKFPDMVHSLKSPVTNVDDPNRFFDFSEPSATHMLTWYSNLGTPASVYRTM
 DGRGVHAYKLNQOQDVNVKQKSOQGISKSLRNKVTQMGKDFNHLNDLYAAIG
 RGNYPKDWLYVKVLSPEALSKLDYNGLDATKVLNVPRDKVTGNTLNRLPENFFLETE
 QSAFAPSNLIPGIEPSDRLLQGRFLFAYATQLYRLGANLFPOLPNRPITSVNNHQN
 GLSNNALQSDGVNVEPSKLNLAEDNQFKAETKLVGTVOQKALSKPRDFYQAGVLY
 RSMNQDSRLTIANLAGDLNKVIDKATWVSFYFRADKEYGSRLAETDTNLSQVK
 NKAMH"
 5962..6429
 /gene="VPA0306"
 5962..6429
 /gene="VPA0306"
 /note="similar to GB:AAF94738.1 (AE004235) percent
 identity 63 in 152 aa"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=11
 /product="AnkB protein"
 /protein_id="BAC61649.1"
 /db_xref="GI:28808472"
 /translation="MKNYFLSSVLSLGSFLSASODKLETOYQSLVELFFDAARI
 GNNVEVTFVSGGFFIDQRNNSEITALMVAAYQGNKDTVRLILLDSGANACLODKRGNT
 ALMGALIKREISIAKDQAECSPOLRNKAGLDLKTFAETIYQSGSEVLSLQNK"

gene	complement(6548. .7201)	
CDS	/gene="VPA0307"	
	complement(6548. .7201)	
	/gene="VPA0307"	
	/note="similar to GB:AAF42465.1 (AE002563) percent identity 47 in 200 aa"	
	/codon_start=1	
	/evidence=not_experimental	
	/transl_table=11	
	/product="putative pyrazinamidase/nicotinamidase"	
	/protein_id="BAC61650.1"	
	/db_xref="GI:28808473"	
	/translation="MTVRVNYQTAVIDNDPEKFGSELCPDELFPVAGALDIVPELLKN HAKGRKLKVRDLHPKAAWDAETPANLPEVGLPNVKNRHCVLGTGGVELLDGL PPVLDYDFQVNGMDPAHPYGIFFHDVADTKTGANEFKCNKIDTVVVGGLADFC VKSVQALDGLFKVIVNLAATRAVLDPDVTDSVAEMKEKRGALFVNADDIIVERFA" complement(7234. .7959)	
gene	/gene="VPA0308"	
CDS	complement(7234. .7959)	
	/gene="VPA0308"	
	/note="similar to GB:AAF96011.1 (AE004352) percent identity 60 in 241 aa"	
	/codon_start=1	
	/evidence=not_experimental	
	/transl_table=11	
	/product="conserved hypothetical protein"	
	/protein_id="BAC61651.1"	
	/db_xref="GI:28808474"	
	/translation="MIVTVDIIPRLSGCADKLEVLIIKRSNPNRPHGVWALPGGF VFNDLTSEGRPADENFEARRRICKIHTIYPRHSEAFIDGDKPRDPEWLSNIT HYALDRNNVVEINNAVPEQKWFPLQALINGEETLAPFDHOKTIEKAWOKURASIE YTSVLLFDKEFLVADIIISAYQFEGIDISRTIKRRLIDSGVLKPTNKVASTNKGK GKPMVYTLTSDSEVTFQNCILRG"	
Query Match	45.68; Score 1031; DB 1; Length 303450;	
Best Local Similarity	67.78; Pred. No. 1.5e-205;	
Matches 1445; Conservative	0; Mismatches 690; Indels 0; Gaps 0;	
Qy	112 GGTAAATGTCCTGTCATGCGGTGTAACACCTCGACCGGTACTTCCAAACAAGATTGG 171	
Db	147685 GGCMAATGCCCGTAATGATGCTGCTCATGACCACCAATAACAGACAGAAAGAACTGG 147626	
Qy	172 TGCGCGGAAGGTTGAACCTGGATATTTTCATCAGCAAGATCGCAATCAGACCGGATG 231	
Db	147625 TGGCCAAAGTCGCTCAACCTCGATATTTTCATCAACACACGCTAAACGGAATCCAATG 147566	
Qy	232 GATCCGGATTCAACTACCGTGAAGAAGTACGCAAGCTCGATTTCGACGCGCTGAAGAAA 291	
Db	147565 CCGAGTGATTTTGATACCAAGAAGAAGTCAAAAACCTCGATTTCCTGCACTCAACAAA 147506	
Qy	292 GATGTCACGCGTTGATGACCGATAGCAAGAGTGGTGGCCGCTGACTGGGGGCACTAC 351	
Db	147505 GACCTATCGCACTAATGACCGATAGCAAGATGGTGGCTGCTGACTGGGACATTAT 147446	
Qy	352 GCGCGTTGATGATCGGTATGGCTGGCACTCGCTGGCACTACCGTATGCTGATGGC 411	
Db	147445 GCGCGTTGATGATGCTATGTCATGSCATGCGGACGGGACTTATCGAATCGGAGATGGT 147386	
Qy	412 CGTGGGGCGGTGGTACCGGAAGCCAGCGCTTTGCAACCGCTCAACTCTCGCGCGGACAAC 471	
Db	147385 CCGCGGGTGGCGGACCGGTAAATTTACGTTTCGCCCGGTAAATTCATCGCCAGATAT 147326	
Qy	472 GTCAGCTTGATTAAGCGCCGCTCTGCTGTGGCCGATCAAGAAGAGTACGGCAACAAA 531	
Db	147325 GCTAATCTGACAAAGCAAGACGCTATCTTTTGCCGATTAAAGAAAATATGTAACCAA 147266	
Qy	532 ATCAGCTGGCGACACCTGATGATTCGGCTGGCGACCGTGGCTTATGAGTCCATGGGCTTA 591	
Db	147265 TTGAGCTGGCGGATCTCATTTGGGTATGCGCGTACGATGGCTTATGAATCCATGGGGTGG 147206	
Qy	592 CCTGCTTACGCTTCTCTTCGCGCGGTGATATTTGGGAACCCGCAAGAAAGATATCTAC 651	
Db	147205 AAAACGTTCCGGCTTCGGGTCGGAAGACATCTGGCCACCCGCAAGAAAGGACATTTAT 147146	

Qy	652 TGGGGTCACGAAAAAGAGTGGCTGGCACCTTCTGACGAACGCTACGCGGACGTGAACAAG 711	
Db	147145 TGGGGCTCAGAAAAAGAAATGGCTGGCACCTTACCAACAAACCTTAACAGCCGCTATTTCAGGT 147086	
Qy	712 CCAGAGACCATGAAAAACCCGCTGGCGGCTGTCCAAATGGGTCTGTATCTATGTGAACCCG 771	
Db	147085 GAACGGGATTTAGAGAAATCCACTCGCGCTGTGATGATGGCTTAATTTATGCAACCCA 147026	
Qy	772 GAAGGTGTTAAAGGCCACCTGATCCCTGAGAACCCGACAGCAGGTACTTTGAACCTTC 831	
Db	147025 GAAGGCGTCGATGGACACCGCGATCCACTCAAAACCCGCGACGCTACGAGTTACGTTTC 146966	
Qy	832 GCGCGTATGCGGATGAACGACGAAAAAACCGAGCCCTCAAGCTGGCGCCACACCGCTC 891	
Db	146965 GCTCGAATGCAATGATGAAGAACCGTTGCCCTTAACCGCTGTGTGCCACACCGTA 146906	
Qy	892 GGTAAATTTGTCACGGTAAATGGCAATGCTCTGCGTTAGCCCTGACCCAAAAGCTCTTGAC 951	
Db	146905 GGTAAAGCGCATGGTAACGGTGTGACGCAATTTAGGCCGACAGCCAGAGGTGCGGAC 146846	
Qy	952 GTTGAANAACGAGGCTTAGGTTGGGCAACCCCAACATGCAAGGCAAGGCAAGCAAGGCC 1011	
Db	146845 ATTACGACCAAGGATTTAGGTTGTTAAACAAGACAAACCGCGCGGTAGCAACAACGCG 146786	
Qy	1012 GTGACCTCGGCTATCGAAGGTGCTTGGACCAACCAACCCCAAGAAATTCGATATGGGCTAT 1071	
Db	146785 GTCAACGATGGGATTTGAAGCGGATGGAACGCAAGCAGCAACGCAATGGGATACGGCTAT 146726	
Qy	1072 TTCACCTGCTGTTCGCTACAAATTTGGAACTGAAAGAGAGTCTCGCGGTGCCACCAT 1131	
Db	146725 TTTTCATCTGCTGTTAAATTTAGGATTTGGAACTGAAAGAAAGCCAGCCGCGCATGSCAA 146666	
Qy	1132 TGGGAACCGATTGACATCAAAAGGAAAAACAACGCCGTTGACGCCAGCACCTCTATT 1191	
Db	146665 TGGGAACCTATGACATTTAAAGAGGAAGAAACACCGGTGCTATCTCTGAAAACCTTAAGGTT 146606	
Qy	1192 CGGCACAACCCGATCATGACCGATGCGGATATGCGGATAAAGGTAAATCCGACTATCGC 1251	
Db	146605 CGGCATAACCCATCATGACCGCGGATATGCGGATGAAGTATGATCGGAGTATCGC 146546	
Qy	1252 GCTATCTGCGAAAAATTCATGGCGGATCTCTGAGTCTTCAAGAAACATTTTCGCGAAGCG 1311	
Db	146545 AAAATCTCAGAACGCTTCCATAGCAGCCGCGCTATTTTCGCCGACACCTTTGCTCGGGCT 146486	
Qy	1312 TGGTTCAAGTCAGCACCGTGAACCTGGGCGCGGAAATCAGCTTACATCGCGCGGAGTG 1371	
Db	146485 TGGTTTAAATTAACGCATAGACATGGGCGCTTAAGGCTCGCTACATCGGCCCTGATGTC 146426	
Qy	1372 CCGCGAGAAGACCTGATTTTGGCAAGACCCGATTTCGCGCAGGTAAACACCGCACTACTCGGAA 1431	
Db	146425 CCGCAAGAAGACCTCATCTGGCAAGATCCGGTTCCTTAACGGAATCAAACTACGATATA 146366	
Qy	1432 GAAGTGGTCAAGCAGAAAAATTCACAAAGTGGCGCTGAGCAATTAAGTAGATGGTCTCCACC 1491	
Db	146365 GACCGGTAAAGCCAAAATCGACGCACTGGGTATCGGTTATGAGTACATCGATCACAAC 146306	
Qy	1492 GCTTTGGGACAGTCCCGTACTTATCGCGTTCCGATATGCGCGCGGTGCTAACGGTGC 1551	
Db	146305 GCTTTGGGACAGTCCGACGCTTCCGTCAGTCTGACAAAGCTGGTGGCGCAACGCGCT 146246	
Qy	1552 CGCATTTGCTTGGCCCCACAGACGAGTGGCAGGCAACGAGCGGAGCCCTGGGGA 1611	
Db	146245 CGTGTTCGCTTAGCACCCACAAAAGACTGCAAGGTAAACGAGCAGAGCGGCTAGCCCGA 146186	
Qy	1612 GTGCTGAGCTCTACGAGCAGATCTCTGCGCACACCGCGGCTAGCATCGCGGACGTCATC 1671	
Db	146185 GTATTACCGGTGTAGAGAACATCGGCAAGATACGGAGCGGCTGCTGATGTGGTC 146126	
Qy	1672 GTTCTGGCGGTAGCGTAGCATCGAGAAAGCCGCGGAAAGCAGCAGGTTACGATGTGGCG 1731	
Db	146125 GTATTACGAGGTAAACGTTCGGTATTGGAACAGCCGCGGCTCGGGGTGTTAACGTCAT 146066	


```

QY 1732 GTTCCCTTCCTGAAAGCGCTGGGATGCGACCGCCGAGATGACCGCAGACTCCTTC 1791
Db 146065 GTCCGCTTCCTCCAGACGCTGGCGACGCAACAGGAGATCGGATGTTGAGTCTTTC 146006
QY 1792 GCACCGCTGAGCGCTGGCGGATGGCTCCGCAACTGGCAGAGAAGAAGATGATGGTG 1851
Db 146005 GAGGTGTACACCGCTTGCACGATGGGTATCGAACTTGGCTCAACAAAACACTACGTTGTA 145946
QY 1852 AAGCCGGAAGAGATGCTGCTGGATCGTGGCAGCTGATGGCTTAACCGCCCGGAAATG 1911
Db 145945 ACACAGAAGAAATGCTGTTGGATCGCACCGCAGTTGATGGGTAACTGCCGCTGAAATG 145886
QY 1912 ACCTGCTGCTGGCGGTATGCGGCTACTGGGCAACCAACTATGGTGGCACAACACAGCG 1971
Db 145885 ACGTGTCTGCTGGAGGATGCGCGTTCCTGGCACAAATCATGCGCGTGAACAAACAGGA 145826
QY 1972 GTATTCACCGATTGTGAAGCGCAGTTGACCAACGACTTTTTTGTGAACCTTGACGATATG 2031
Db 145825 GTGTTACACAGCGAGTCGCGCAACTTACCAACGATTTTTTCATTAACTTAACCGACATG 145766
QY 2032 GGGAAACGTCGGAAGCGGTAGGTAGCAACGCTACGAAATCGCGGACCGAAGACCGGT 2091
Db 145765 AAGTACACGTTGGGACCGCTTGGCGAAACGCTTTACGAAATCGGTGCGCTCCAAA 145706
QY 2092 GCCGTGAAGTGGACCGCTCGCGGTGATGCTGTTGTTGTTCCAACTCGCTACTCGC 2151
Db 145705 GATGTAATATGGACGAGCGACCGCGTGCATCTGTCTTCGGCTCAAACTCAATACTCAGA 145646
QY 2152 TCTTACGCAAGTGTACGCGCGAGGAGATAAACGGCGAGAAGTTCGTCAGAGACTTCGTC 2211
Db 145645 GCCTATGCCGAACGTATGCGCAAGATGACACCGCAGGAAATTCGTTGAAGACTTTGTC 145586
QY 2212 GCCGCTGGACAAAGTGTATGAAGCGCGACCGGTTT 2246
Db 145585 GCCGCTGGACAAAGTGTATGAAGCGCGACCGATT 145551

RESULT 11
AF078110
LOCUS AF078110 3865 bp DNA linear BCT 17-OCT-1998
DEFINITION Legionella pneumophila catalase-peroxidase KatB (katB) gene,
complete cds.
ACCESSION AF078110
VERSION AF078110.1 GI:3764089
KEYWORDS Legionella pneumophila
SOURCE Legionella pneumophila
ORGANISM Legionella pneumophila
Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
Legionellaceae; Legionella.
1 (bases 1 to 3865)
Bandyopadhyay,P. and Steelman,H.M.
Legionella pneumophila catalase-peroxidases: cloning of the katB
gene and studies of KatB function
J. Bacteriol. 180 (20), 5369-5374 (1998)
98440428
MEDLINE
PUBMED 9765568
REFERENCE 2 (bases 1 to 3865)
Bandyopadhyay,P. and Steelman,H.M.
Direct Submission
Submitted (14-JUL-1998) Biochemistry, Albert Einstein College of
Medicine, 1300 Morris Park Avenue, New York, NY 10461, USA
FEATURES
source
1. .3865
/organism="Legionella pneumophila"
/mol_type="genomic DNA"
/strain="JR32"
/db_xref="taxon:446"
882..3047
/gene="katB"
882..3047
/gene="katB"
/codon_start=1
/transl_table=11

```

```

/product="catalase-peroxidase katB"
/protein_id="AAC64361.1"
/db_xref="GI:3764090"
/translation="MDGKVGSTTIGCPVINGMTSTGTSNTAWPNALNLDLHOHDT
KTPMEKDPNREBEVKLDFEALKKDLHALMTDSQAWPADWGHYGLMLRMSHIAAG
SYRVADRGAGCTGNQRPAPLNSWPDNVNLDKARLLWPILKKYGNKISWADLIVLAG
TIAESMLGTLFGFGREDIWHPEKDVYWGSEQEWLGAKRIDYKSKRESLENPLAAVQ
MGLIYYNPEGVNGOPDPLRTAQDVRTVFGRMAMNDEETALTAGGHTVGVCHGNAGK
WELKPSAGAWQEPYIHKEDKPDVEDPARIHNPIMTDADMAKMDPVYKIAERF
YQDPDYFAEVFAWPKLTHRDMPKTRYIGPDVPKEDLWQDPVPAGNAYDIAAK
AKTAASNLITGEWVSTAWDSARTFSGDKRGGAANGARILKPKQDWZGNRPQRLTKVL
QILDIATDTGSGVADVILLAGNVGIEKAARAGFDIIVFPARGGATDDMTDAESF
DVLEPHDGYRNMLKTYDVRPEELMDLDTLMGLTAHEMTVLGGRLVLTGTHNNHTQ
YGVFTDRVGLATNDFVNLTDMAVNWIPSKDNLNLEIRDRKAGNIKWTATRVDLVFGSN
SILRSVAEVTAQDDNKGKTFQDFVAAATKVMNADRFDLA"
BASE COUNT 1150 a 809 c 858 g 1048 t
ORIGIN
Query Match 45.6%; Score 1030.4; DB 1; Length 3865;
Best Local Similarity 68.0%; Pred. No. 2.6e-205;
Matches 1455; Conservative 0; Mismatches 676; Indels 9; Gaps 1;
QY 118 TGTCTCTCATGACGCGTGTAAACACCTCGACCGGTACTTCCAACAAGATTGCTGGCGG 177
Db 915 TGTCCGCTGATACATGCGGTATGACGCTACAGGAACATCCACACGGCATGCTGGCCA 974
QY 178 GAAGGGTGAACCTTGGATATTTTCATCAGCAAGATCGCAATCAGACCCGATGGATCGG 237
Db 975 AATGCCTTGAACCTGGATATATTCATCAGCATCAGTACACCAAGCAATCCANTCGAANA 1034
QY 238 GATTTCAACTACCGTGAAGAGTACGCAAGTCGATTTCGACGCGCTGAAGAAGATGTC 297
Db 1035 GATTTCAACTACCGTGAAGAGTCAAAAACTGGATTTCGAGGCACTCAAAAAAGATTG 1094
QY 298 CACGCGTTGATGACCGATAGCAAGAGTGGTGGCCCTGACTGCTGGGCGCACTAGCGCGT 357
Db 1095 CATGCTTAATGATGACGATGACGCGTGGTGGCCCTGCTGATTGGGACACATACGGTGGT 1154
QY 358 TTGATGATCGGTATGGCTTGGCACTCGGCTGGCACTACCTACCTATGTTGATGGCGCTGGG 417
Db 1155 CTATGATTCGATGCTCTCGCATGCTGCTGCTTCTTCTATCGCTGGCTGATGGCGTGGC 1214
QY 418 GCGCGTGGTACCGGAAGCAGCGCTTTGACCGCTCAACTCTCTGGCGGACAAGCTCAGC 477
Db 1215 GGTGCAAGTACTGGCAATCAGCGTTTGGCTCATTTGAATTCCTGGCTGACAAATGTGAAC 1274
QY 478 CTGCATAAAGCGCGCTGCTGCTGGCCGATCAAGAAGATGACGCAACAAATCAGC 537
Db 1275 CTGGATAAGCAGCCTGCTGTTTATGGCCGATCAAGAAAAAATACGGCAACAAATCAGT 1334
QY 538 TGGCAGACCTGATGATTCTGGCTGGCACCCTGGCTTATGATGATCCATGGGCTTACCTGCT 597
Db 1335 TGGCAGACCTGATGATTCTGGCCGCTGATTCCTTATGATCAATGGGACTTAAACA 1394
QY 598 TACGGCTTCTTTTCGGCGCGCTCGATATTTGGGAACCCGAAAAAGATATCTACTGGGGT 657
Db 1395 TTCGGTTTTGGTTTTGGCCGTGAAGACATTTGGCATCTGAAAAAGATGTTTACTGGGGT 1454
QY 658 GACCAAAAAGAGTGGCTGGCACCCTTCACCAACGCTACGCGGACGTGAACAGCCAGAG 717
Db 1455 TCAGAAACAAGAAATGGTTG-----GGGCCCAAGCGTTTACGATGGCAAAAGCCGAG 1505
QY 718 ACCATGGAAAAACCGCTGGCGCTGTCCAAATGGGTCTGATCTATGTGAACCCGGAAGGT 777
Db 1506 TCTCTGGAAAAACCGCTGGCGGAGTACAAATGGGATTAATTTATGTAATATCCCGAAGCA 1565
QY 778 GTTAAACGCCACCTGATCCCTGAGAACCCGACAGCAGGATCTTGAACACCTTCGCCCGT 837
Db 1566 GTGAATGGACAACCTGATCCCTTTCGACACGCGCAGGATGTTCTGTGTAACCTTCGGACGG 1625
QY 838 ATGGCGATGAACGACGAAAAAACCCGACCCCTCACAGCTGGCGGCGACACCGCTGTAAT 897

```

Db 1626 ATGCCAATGAACGACGAGGAAACCGTTGCTCTGACTGCCGGAGGACATAGTGTGGCAA 1685
 QY 898 TGTACAGGTAATGGCAATGCTCTGGTGTAGCCCTGACCCAAAGCCTCTGACGTTGMA 957
 Db 1686 TGTATGGCAATGGCAATGCAAAACCTTTGGGCCCCCAATCCTGAGCGGCTATGTGTGAG 1745
 QY 958 AACCAAGGCTTAGGTTGGGGCAACCCCAACATGCGAGGGCAAGGCAAGCAACGCGGTGACC 1017
 Db 1746 GACCAAGGGCTTGCTGTGATTACAAACCAACCCAGAGGATAGCGGCTAATACTGTTTCA 1805
 QY 1018 TCGGGTATCAAGGCTCTGGACCAACCAACCCAGCAATTCGATATGGCTATTTCGCAC 1077
 Db 1806 AGCGGATTTGAAGCGCGATGGAACAACACCCCTACCCGAGGACATGGCTACTTCTAT 1865
 QY 1078 CTGCTGTGCGGTACAAATTTGGGAACCTGAAAGAGTCTTGCCTGCGCCACCACTTTGGGAA 1137
 Db 1866 TTATTGCTGAATACGATTGGGAACCTGAAAGAGTCTTGCCTGCGCCGCTGGCAATGGGAG 1925
 QY 1138 CCGATTGACATCAAAAGAAACGCGGTTGACGCGAGCGACGCCCTCTATTTGCCAC 1197
 Db 1926 CCCATCCATATTAAAGGAAGAAGATAAACCAAGTAGATGTTGAAGACCTGCGATCCGGCAC 1985
 QY 1198 AACCCGATCATGACCGATCGGATATGGGATAAAGGTAAATCCGACCTATCCGCTATC 1257
 Db 1986 AACCCGATCATGACTGATGCTGACATGGCAATGAAATGATGCCCTTTATCGTAAAT 2045
 QY 1258 TGGGAAAAATTCATGCCGATCTCGATCTTCAAGAAACCTTTCCGGAAGCGGTGCTC 1317
 Db 2046 GCAGAGCGCTTTTACCAAGATCCGACTATTTTGCAGAAAGTTTTCGCTGAGCGCTGCTC 2105
 QY 1318 AAGCTGAGGACCGTGACCTGGCCCGCAATACGTTACATCGCCCGGGAAGTGGCGCA 1377
 Db 2106 AACTGACTCATCGGATATGGGCGCAAGACGCGCTACATTTGCGCCTGATGTTCTCTAAA 2165
 QY 1378 GAAGACTGATTGGCAAGCCGATTCGGCGAGGTAAACCGACTACTCGGAAGAGTG 1437
 Db 2166 GAAGATTTAATTTGGCAGATCCGTCCTGCTGTTGTAACAGGCAATGATATGCTGCC 2225
 QY 1438 GTCAAGCAGAAATTCACAAAGTGGCCTGAGCATTAGTAGTGTCTCCACCGCTGG 1497
 Db 2226 GCCAAAGCAAAATTTGCGCCAGCAATCTGACCATCGGTGAATTTTCTACTGCATGG 2285
 QY 1498 GACAGTCCCGTACTTTATCGCGTTCCGATATCGCGGGGGTGTACAGGTGCCCGCAT 1557
 Db 2286 GACAGCGCGCACTTTTCGCGGCTCAGATAAGCGCGTGGTGCAGAACGAGCAGCATATA 2345
 QY 1558 CGTTTGGCCCAACAGATGGCAGGCAACGAGCGGAGCGCTGCGGAAGTGGTG 1617
 Db 2346 CGACTAAAGCCACAAAGACTGGGAAGGCAATGAGCCACAGCCCTGACTAAAGTCTTG 2405
 QY 1618 AGGCTCTACAGCAGATCTCTCCGACACCGGCGCTAGCATCGCGGACGTGATCGTTCTG 1677
 Db 2406 CAAATTTCTTGAAGACATCGCTACTGATACAGGCGCAAGCGTGGCAGATGTGATTATCT 2465
 QY 1678 GCGGGTAGGCTAGGATCAGAAAGCCGGAAGCAGCAGGTTACGATGTGCGCTTCCC 1737
 Db 2466 GCGGAATATGCGGTATCAGAAAGGCTGCCAAGGCAAGCTGGCTTTGACATTTATCGTCCC 2525
 QY 1738 TTCTCTAAAGGCGGTGGCATGCGACCGCGAGATGACCGAGCAGACTCTTCGCACCG 1797
 Db 2526 TTGTCTCTCGCGCTGGGATGGAACCGATGACATGACCGATGCGAATCCCTTTGAGT 2585
 QY 1798 CTGAGCCGCTGGCCGATCGCTTCCGCAACTGGCAGAAAGAAAGTATGTGTTGAAGCCG 1857
 Db 2586 CTGGAACCCGCTCATGATGGCTATCGCAACTGGCTTTAAGAAAGACATATGATGTCGCTCT 2645
 QY 1858 GAAGAGATCGCTGCTGATGCTGGCGAGCTGATGGGCTTAACCGCGCCGGAATGACCGTG 1917
 Db 2646 GAAGAGCTCATGCTAGATGCGCACTCAGCTCATGGGACTGACAGCCCATGAGATGACAGTC 2705
 QY 1918 CTGCTGGGCGGTATGCGGCTACTTGGCCACCACTATGTTGGCCACCAACACCGCGTATTC 1977
 Db 2706 CTGTGTTGGGGTCTGCGCGTGTGGTACTAATACAAATACCCAAATATGCGGTATTC 2765

QY 1978 ACCGATTGTGAAGCGCCAGTTGACCAACGACTTTTTTTGTGAACCTGACCATATGGGGAAC 2037
 Db 2766 ACTGATCGGTTGGTCCCTGACTAATGACTTTTTTTGTCACTTGACCATATGGCCAAT 2825
 QY 2038 AGCTGGAAGCGGTAGTGTAGCAACGCTTACGAATCCGCGACGCAAGACCGGTGCGCGTG 2097
 Db 2826 GTATGATATCCCAAGCAAGACAAATTTGTACGAATTCGAGACCGTAAAGCGGTAAACATC 2885
 QY 2098 AAGTGACCGCTCGCGGGTGGATCTGTATTTGGTTCCAACTCGCTACTGCGCTTTCAC 2157
 Db 2886 AAGTGACAGCGACTCGTGTAGATCTGTTTGGCTCTAATTCGATTCGCGCTCTAT 2945
 QY 2158 GCAGAAGTGTAGCCCGACGACGATACCGGAGAACTTCGTCAGAGACTTCGTCGCGCG 2217
 Db 2946 GCTGAAGTCTATGCCAGGATGATACAGGGGAAATTCATACAGGACITTTGTGAGCA 3005
 QY 2218 TGGACCAAGTGTAGAACCGCCGACCGCTTTTCGAGCTGCGCT 2257
 Db 3006 TGGACAAAGTTATGAATGCTGATCGCTTTGATTTGGCAT 3045

RESULT 12

SYOGP
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 FEATURES
 source

SYOGP
 Synchococcus sp. DNA 2430 bp DNA linear BCT 12-FEB-1999
 D61378
 D61378.1 GI:1550616
 catalase-peroxidase.
 Synchococcus sp.
 Bacteria; Cyanobacteria; Chroococcales; Synchococcus.
 Mutsaers, M., Ishikawa, T., Takeda, T. and Shigeoka, S.
 The catalase-peroxidase of *Synchococcus* PCC 7942: purification,
 nucleotide sequence analysis and expression in *Escherichia coli*
 Biochem. J. 316 (Pt 1), 251-257 (1996)
 96235199
 8645214
 2 (bases 1 to 2430)
 Shigeoka, S.
 Direct Submission
 Submitted (26-JUN-1995) Shigeru Shigeoka, Faculty of Agriculture,
 Kinki University, Department of Food and Nutrition; 3327-204
 Naka-machi, Nara, Nara 631, Japan (Tel:0742-43-1511(ex.3416),
 Fax:0742-43-1155)
 On Sep 19, 1996 this sequence version replaced gi:1359463.
 Sequence updated (12-Sep-1996) by:Shigeru Shigeoka.

1. .2430
 /organism="Synchococcus sp."
 /mol_type="genomic DNA"
 /strain="PCC7942"
 /db_xref="taxon:1131"
 167. .171
 181. .2343
 /codon_start=1
 /transl_table=11
 /product="catalase-peroxidase"
 /protein_id="BA009601.1"
 /db_xref="GI:1359464"

RBS
 CDS

/translation="MTATGGKCPVMHGGATTVNISTAEWPKALNDILSOHDKRTNP
 MGFDFNYEVEQKLDALQDLQALMTDSQWPDWGHYGLMRLWVHAAGTVRIA
 DGRGAGTGNORFAPLNSWPNDNLKARLLWP IKQVGNKLSWADIAVAGTAYE
 SNGLTFGAFGRDINHPKDIYNGPEKFPSTNPNSRVTGDELENPLAAYTMG
 LIYNPEVDGNPDLKTAHDVRYTFARMAMDEETVALTAGGHTVGKCHGNGNALL
 ELKPNAGAWQHEP INPREEDLPVDEPSTRNLVMTDADNMDKNDPEYKISRFP
 QDPAYFADVFARAFKLRHMDRQKARYIGPDVPOEDLIWODIPAGNRNVDQAVK
 RTAAGSLTSELSTANDSARTYRNSDKGGANGARIRLAPKQKNEGNEPDLKPLA
 VLEGISAATGATADVILVAGNVGVEOKARAGVEIPLFPAGGADTAEGQDTSEFA
 VLEPHDAITGSSRTMRQLKNCCLLATQLGLTAPETVLIIGLRLVNHGTTKH
 VVFTDREGVLTNDFVFNLTDMNLMKPAKGNLYETCDRRTNOVKWATRVLDVFGSNS

BASE COUNT	611 a	588 c	697 g	534 t	ILRAYSELVYAQDDNKEKFFVDFVFAVNTKVMNADRFILD*
ORIGIN					
Query Match	45.1%	Score 1019.2;	DB 1;	Length 2430;	
Best Local Similarity	68.2%	Pred. No. 6e-203;			
Matches 1474;	Conservative	0;	Mismatches 678;	Indels 8;	Gaps 4;
Qy	99	GAGAGCTATGAGCGGTAAATGCTGCTGATCGACGGTGGTAACACCTTCGACCGGTACTTC	158		
Db	183	GACAGCAACCCAGGTAATGTCGGGTCAATGACGGGAGCAACACCTTAATATTC	242		
Qy	159	CACAAAGATGGTGGCGGAAGGGTTGAACCTGGATATATTTGCATGACGAAGATCGCAA	218		
Db	243	GACTGCTGAGTGGTGGCAAGGCACTCAACCTGGATATTTTGAAGCCAGCAGCATGCGAA	302		
Qy	219	ATCAGACCCGATGATCCGATTCACCTACCGTACGAGAGTACGCAAGCTCCATTTGCA	278		
Db	303	GACCAACCAATGGGCGACACTTCAACTATCAGAGAAAGTCCAGAACTGGAT--GC	359		
Qy	279	CGGCTGAAGAAAGATCTCCACCGTTGATGACCGATAGCAAGAGTGGTGGCGCGGTGA	338		
Db	360	TGCGCTCAAGCAAGATTACAGGCGCTGATGACCGATAGCCAGTGGTGGCGGCGAGA	419		
Qy	339	CTGGGGCACTACGGCGGTTGATGATPCGTATGGCTTGGCATTCGCGTGGCACCTACCG	398		
Db	420	CTGGGTCACCTACGGCGCTGATGATTTCGCCCTCACTTGGCACGGGCGGCACTACCG	479		
Qy	399	TATGCTGATGGCGTGGGCGGTGTACCGAGCCAGCGCGCTTGGCACCGCTCAACTC	458		
Db	480	AATTGGCCGATGGTGGCGGTGGTACGACAGGGGAACAGCGCTTGTCTCCCTCAATTC	539		
Qy	459	CTGGCGGACACGTCAGCTGGATAAAGCGCGCGTCTGCTGGCCGATCAAGAGAA	518		
Db	540	TTGGCCAGACACAAATTTAGACAAGCGCGTGGCTTGGCCGATCAAGCAAAA	599		
Qy	519	GTAGGCAACAAATACAGCTGGGACAGCTGATGATTCGGCTGGCACCGTGGCTTATGA	578		
Db	600	GTAGGCAACAAATTTGATTTGGGAGATTTAATTTGCCCTATGCGGCGACGATCGCTACGA	659		
Qy	579	GTCCATGGGCTTACCTGCTTACGGCTCTCTTTGGCGCGGTGATATTTGGAAACCGGA	638		
Db	660	ATCGATGGGCTTAAACCTTTGGTTTGGCTTTGGACGAGAGATATTTGGGATCCTGA	719		
Qy	639	AAAAGATATCTACTGGGGTGACGAAAAGTAGTGGCTGGCACCTTCTGACGAACGCTACGG	698		
Db	720	GAAAGATATCTACTGGGGCTTGAGAGGATGGTTCCCCCAAGCACCAATCCCAACAG	779		
Qy	699	CGAGTGAAACGCCAGAGACCAATGGAACACCGCGTGGCGGTGTGCAATGGGTCTGAT	758		
Db	780	TCGCTATACGGGGGATCGCAACTTGAATAATCCGCTAGCAGCGGTGACATGGGCTGAT	839		
Qy	759	CTATGTCAACCCGGAAGTGTTAACGGCCACCTGATCCGCTGAGACCCGACAGAGGT	818		
Db	840	TTAGCTCAACCCGGAAGCGGTGGATGGCAATCTGATCCGCTCAAAACCGCCCATGACGT	899		
Qy	819	ACTTGAACCTTTCGCCGATGCGATGAAACGAGCAAGAAAACCGCACCGCTCACAGCTGG	878		
Db	900	CGCGCTCACCTTTGGCGGGATGGCGATGACGATGAGGAACGGTGGCGCTTAAGTCTGG	959		
Qy	879	CGGCCACACCGTCCGTAATTTGACGGAATATGGCAATGCCCTCTGCGTTAGCCCTGACCC	938		
Db	960	TGGACACACCGTTGGCAATGTCTATGGCAATGGCAATGGCTTGTGCTAGGACCCGAACC	1019		
Qy	939	AAAAGCTCTGACGCTTGAACACAGGCGTTAGTTGGGGCAACCCCAACATGACGGCAA	998		
Db	1020	GGAAGGGCGGATGTGAAGATCAAGGCTTTGGGCTGGATCAATAAAAACCCAGAGCGGTAT	1079		
Qy	999	GGCAAGCAACCGCTGACCTCGGATATGCAAGGTGCTTTGGACCAACCAACCCACGAAAT	1058		
Db	1080	TGGTCGCAACGCTGTACCAAGTGGGCTGGAGAGGGGCTTGGACACCCCAACCGGACTAAT	1139		
Qy	1059	CGATATGGGCTATTTCGACCT--GCTGTTCCGCTACAATTTGGGAACTGAAAAGAGTCC	1115		

Db	2219	CAAAATTCGATTCGCGACCTACTCAGAGCTCTATGCACAAGACGACACACAAGAGAAGT	2278
QY	2195	TGCTCAGAGACTTCCTCGCGCTCGACCAAGATGATGAAGCCGACGTTTCGACGTGC	2254
Db	2279	TTGTGCGAGACTTTCGCTGCTGCTGCGACGAAGGTGATGAATGCCGATCGCTTTGATCTGG	2338
RESULT 13	AE015873/c		
LOCUS	AE015873.1	12293 bp	DNA linear BCT 02-DEC-2002
DEFINITION	Shewanella oneidensis MR-1	section 422 of 457	of the complete genome
ACCESSION	AE015873	AE014299	
VERSION	AE015873.1	GI:24350840	
KEYWORDS			
SOURCE	Shewanella oneidensis MR-1		
ORGANISM	Shewanella oneidensis MR-1		
REFERENCE			
AUTHORS	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Alteromonadaceae; Shewanella.		
	1 (bases 1 to 12293)		
	Heidelberg, J., Paulsen, I., Nelson, K., Gaidos, E., Nelson, W., Read, T., Eisen, J., Seshadri, R., Ward, N., Methe, B., Clayton, R., Meyer, T., Tsapin, A., Scott, J., Beanan, M., Brinkac, L., Daugherty, S., DeBoy, R., Dodson, R., Durkin, A., Haft, D., Kolonay, J., Madupu, R., Peterson, J., Umayam, L., White, O., Wolf, A., Vamathevan, J., Weidman, J., Impraim, M., Lee, K., Berry, K., Lee, C., Mueller, J., Khouri, H., Gill, J., Utterback, T., McDonald, L., Feldblyum, T., Smith, H., Venter, J., Nealon, K. and Fraser, C.		
TITLE	Genome sequence of the dissimilatory metal ion-reducing bacterium Shewanella oneidensis		
JOURNAL	Nat. Biotechnol. 20 (11), 1118-1123 (2002)		
MEDLINE	12397686		
PUBMED	12368813		
REFERENCE			
AUTHORS	2 (bases 1 to 12293)		
	Heidelberg, J., Paulsen, I., Nelson, K., Gaidos, E., Nelson, W., Read, T., Eisen, J., Seshadri, R., Ward, N., Methe, B., Clayton, R., Meyer, T., Tsapin, A., Scott, J., Beanan, M., Brinkac, L., Daugherty, S., DeBoy, R., Dodson, R., Durkin, A., Haft, D., Kolonay, J., Madupu, R., Peterson, J., Umayam, L., White, O., Wolf, A., Vamathevan, J., Weidman, J., Impraim, M., Lee, K., Berry, K., Lee, C., Mueller, J., Khouri, H., Gill, J., Utterback, T., McDonald, L., Feldblyum, T., Smith, H., Venter, J., Nealon, K. and Fraser, C.		
TITLE	Direct Submission		
JOURNAL	Submitted (12-SEP-2002) The Institute for Genomic Research, 9712		
MEDLINE	Medical Center Dr., Rockville, MD 20850, USA		
PUBMED	Location/Qualifiers		
REFERENCE			
AUTHORS	1. .12293		
	/organism="Shewanella oneidensis MR-1"		
	/mol_type="genomic DNA"		
	/strain="MR-1"		
	/db_xref="taxon:211586"		
	288. .1022		
	/locus_tag="S04391"		
	288. .1022		
	/locus_tag="S04391"		
	/note="identified by Glimmer2; putative"		
	/codon_start=1		
	/transl_table=11		
	/product="hypothetical protein"		
	/protein_id="AA057359.1"		
	/db_xref="GI:24350841"		
	/translation="MRQPEGNIRLRYMDFSKFVALISSKSLFLNRVDRFTDPYEGH FSSLNRRINSIYEAQSGSIDPHKVANAISDFNSFROWYVNCWHANKYESAAMW DLYAKTEBAIAETDYNLSLAVLPDQAOLGLVEYIDPDTMPPENNTVTPVHKRSF EHERKRVIIVEIIPRAKSGVGFDEAVNDKHSGLSIDLNQLKSIHVSPTAPKWLVD LTNVTSYFGIDAHKQSDIUKAPIK"		
	complement(1151. .1264)		
	/locus_tag="S04392"		
	complement(1151. .1264)		
	/locus_tag="S04392"		
	/note="identified by Glimmer2; putative"		
	/codon_start=1		
	/transl_table=11		
	/product="hypothetical protein"		
	/protein_id="AA057360.1"		
	/db_xref="GI:24350842"		
	/translation="MPCGNSGMYSLGSGVIGOMLNEKSYACPIILFLWL"		
	1691. .2146		
	/locus_tag="S04393"		
	1691. .2146		
	/locus_tag="S04393"		
	/note="similar to SP:P33599, and PID:397900; identified by sequence similarity; putative"		
	/codon_start=1		
	/transl_table=11		
	/product="acetyltransferase, GNAT family"		
	/protein_id="AA057361.1"		
	/db_xref="GI:24350843"		
	/translation="MKITLDLKGPEIAALQEHLDMMRATSPESHALDLNGLRQP NIRFTWLDDRNLAGCGALKWLDHAHEIKSMRTAATYKQGVASQILQHLINDAKRA GQVRLSTGSMTFQFARSLYAKFGFELCGPFADYVLDPSVFMTKNI"		
	2279. .2605		
	/gene="pspE-2"		
	2279. .2605		
	/locus_tag="S04394"		
	2279. .2605		
	/gene="pspE-2"		
	/locus_tag="S04394"		
	/note="similar to GB:X58480, and PID:36070; identified by sequence similarity; putative"		
	/codon_start=1		
	/transl_table=11		
	/product="phage shock protein E"		
	/protein_id="AA057361.1"		
	/db_xref="GI:24350844"		
	/translation="MFQSLAKLSGSSNGEKWLEQGARVIDRSPREFASGHLPQ AINVPLPTLDQWLHQVENKQHPFVLYCGAGIRAQGCODILRANGFCVYVNGSLKLV CCQNP"		
	2645. .2809		
	/locus_tag="S04395"		
	2645. .2809		
	/locus_tag="S04395"		
	/note="identified by Glimmer2; putative"		
	/codon_start=1		
	/transl_table=11		
	/product="hypothetical protein"		
	/protein_id="AA057362.1"		
	/db_xref="GI:24350845"		
	/translation="MTTSCPAPOAEKPAELKSARQARYAMSRNGQASVKVYIETS AMSLGKICE"		
	complement(2871. .3467)		
	/gene="acpB"		
	complement(2871. .3467)		
	/gene="acpD"		
	/locus_tag="S04396"		
	/note="similar to SP:P41407, SP:P33548, PID:155324, and PID:155328; identified by sequence similarity; putative"		
	/codon_start=1		
	/transl_table=11		
	/product="acyl carrier protein phosphodiesterase"		
	/protein_id="AA057363.1"		
	/db_xref="GI:24350846"		
	/translation="MSKVLVKSSILGGYSQSALLVDYLGKWEKOGATTTVRDLQAG DVLPMVGGETASGLRGGAEL		

/db_xref="GI:24350847"
/translation="MDLLNRLQTHWSTIPVSEFMQIAPLSFTDGLSVSAPLAPNI
NLHRTFAGSYITMTLGTGWMVLAQQLNLDVGDIVLADAHIRYLAPVTSAPVKVR
WPDNLSPGLGRKAKVKLEVLQFCGDKLCAQFDGLVSVPKM"
complement(4065..4502)
/locus_tag="S04398"
complement(4065..4502)
/locus_tag="S04398"
/note="similar to GB:X57352, SP:Q01628, SP:Q01629, and
PID:23398; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein TIGR00256"
/protein_id="AA057365.1"
/db_xref="GI:24350848"
/translation="MTALIQVRASVYVDNQTIGATDKLLVLLGVREDNRKMEK
LATKVSRYVFSDEKMNLNLTQAGSLVWSQFTLAADTERGLRSPFCAGTPEGA
LGLVEEFAFCRHKGVNTEGTQGAADMKVELVNDGPVTFHLQV"
complement(4518..5132)
/locus_tag="S04399"
complement(4518..5132)
/locus_tag="S04399"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA057366.1"
/db_xref="GI:24350849"
/translation="MKLAIILITGLTACAATPSASVAGELAKOPSVSAASTDARA
PVAPKVLQAGMNEVADALYOALLNANKVGPDKVAVQFNDFLESMNNA
GRILNLRNVAHPQLOASPDPVAVGELNKLNFVLRQGVAVIQIGVITFAD
KVEELRRLMWTQALQIAQSLPDGVSEYLRPIPLMPVKS"
complement(5129..6085)
/locus_tag="S04400"
complement(5129..6085)
/locus_tag="S04400"
/note="similar to GP:1710151, GB:X62583, and PID:42775;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="proline iminopeptidase, putative"
/protein_id="AA057367.1"
/db_xref="GI:24350850"
/translation="MLTDTVQVAPFIRRDWMMGSGQQLHLAQYNPQGIPLLYLHGG
PGASVSELSLENPEHYWILLIDQKAGSLPSGELEHNLNGLICDIEAIRLGI
ERWCLAGSGATLALYGLFPNRIAVQIWMALFIPSKCEIWLTPSGAAQLYPOA
HREFAAPSGIDLFTHYQLGNAQDEATREFARRWIOVELTLGAVPIALPRLLSAQ
QLAAQLTELHAQNGYFNMFSLQRTVSQVARTLLQGTQDVCVCPARLLATFLTKVG
NHRTEIHSIVDGGHLSNLSLAVTLEIQAMWIKRQELT"
complement(6289..7176)
/gene="rbn"
/locus_tag="S04401"
complement(6289..7176)
/gene="rbn"
/locus_tag="S04401"
/note="similar to GB:J03161, SP:PI1831, and PID:338480;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ribonuclease Bn"
/protein_id="AA057368.1"
/db_xref="GI:24350851"
/translation="MTRKIDLAQIRVILLGIWHLRLRORLVEDQINIRAGHLAYVT

Query Match
Best Local Similarity 42.1%; Score 951.2; DB 1; Length 12293;
Matches 1423; Conservative 65.7%; Pred. No. 9.5e-189;
Mismatches 728; Indels 15; Gaps 2;
112 GGTAAATGTCCTGTCATGCAGCGGTGGTAACACCTCGACCGGTACTTCCAAACAAGATGG 171
11966 GGTAAATGCCCGTGATGCATGCGGCAATTCCTCCGCTTGCCACTAACAAATATGACTGG 11907
172 TGGCCGGAGGGTTGACCTGGATATTTTGCATCAGCAGATCGCAATATCAGACCGCGATG 231

11906 TGGCCTTAAGCTCTTAACCTCGATATCTTGCACCAACACCATCAGAAACGGATCCGATG 11847
232 GATCCGCGATTTCAACTACCGTGAAGAAGTACGCAAGCTCGATTTTCGACGGCTCAAGAAA 291
11846 GACCTAACCTTAGTTATCGTGATGCATTCAACAGTTTAGATCTCGACGAGTAGACAG 11787
292 GATGTCACCGCTTGTGATGACCGATAGCAAGAGTGGTGGCCCGCTGACTGSGGGGACTAC 351
11786 GACTTACATTCCTTATGACCGATAGCAAGATTTGGTGGCCAGCGGATTTGGGGCATTTAT 11727
352 GCGCGTTTGTATGATCCGATATGGCTTGGCACTCCCTGGCACCTACCGTATGCTGTATGGC 411
11726 GCGCGTTTGTATGATCCGATATGGCATGGCACTCCGACGAACTATCGGATTTGCAGATGGC 11667
412 CGTGGCGCGCTGTGTACCGAAGCAGCGCTTTCACCGCTCAACTCTCTGCGCGGACAAC 471
11666 CGTGGCGCTGGGGGACGGGAATCAACGCTTTCGCGCGCTCAATAGCTGGCAGATAAC 11607
472 GTCAGCTTGTGATAAAGCGCGCGCTCTGCTGTGGCCGATCAAGAGAAGTAGGCAACAAA 531
11606 GCCAACCTTGATAAAGCGCGCTGCTTACTTTGGCCGATTAAGAAAAAATACGGCAATAAA 11547
532 ATCAGCTGGCGACGCTGATGATTCTGGCTGGCACCGCTGGCTTATGAGTCCATGGCTTA 591
11546 CTCTCATGGCGCGGATTTGATTTATCTGCTGGCAACGCTGGCTATGAATCTATGGGATTA 11487
592 CCTCTTACGGCTCTCTCTTTCGGCGCGCTCGCATTTTGGGAACCGCAAAAAGATATCTAC 651
11486 AAAACCTTTGGCTTTGGCGGTGGCTGGCCGATATCTGGCATCTCTGAAAAGACATTTAT 11427
652 TGGGGTGACGAAAAGAGTGGCTGGCACCTTCTGACGAACGCTACGGCGAGCTGAACAAAG 711
11426 TGGGGCTCAGAAAAACAATGGCTCGCGCGGATGATAATCCGACAGCTGTTACTCTGGT 11367
712 CCAGAGACCATGAAAAACCGCTGGCGGTGCTTCAAAATGGGTCTGATCTATGTGAACCG 771
11366 GCTCGACCTTGAAAAATCCCTTGGCGCGCTTAATGATGGGGCTGATTTATGTGAATCCC 11307
772 GAAGGTGTTAACGGCCACCTTGATCCGCTGAGAACCGCACAGCAGGTACTTTGAAACCTTC 831
11306 GAAGGATTTGATGTTAGCCCGATCCACTGCGCACCGCCCAAGATGCGGTATCACCTTT 11247
832 GCCCGTATGGCGATGAACGACGAAAAAACCCGACCGCTTCACAGCTGGCGGCACACCGCTC 891
11246 GCGCGGATGGCGATGATGAGAAGAACCGCTCGCTTAACCTGGGGTGGCCACACTGTA 11187
892 GGTAAATGTCACGCTAATGGCAATGCTCTCGGTAGCCCTGACCCAAAGACCTCTGAC 951
11186 GGTAAATGCCCATGTTGTTAAAGCTCAAGACCTTGGCCCGCAACAGAGGGTGAGGAT 11127
952 GTTGAAACCAAGGCTTAGGTTGGGCAACCCCAACATGCAAGGCAAGCAACGCC 1011
11126 GTTGAAGCGCAAGGTTTAGTTGTTTAAACAAAAAGCCCGCTGGCTTGGCTGTGATGCT 11067
1012 GTGACCTCGGGTATCGAAGTGTGTTGGACCAACCAACCCAGAAATTCGATATGGGCTAT 1071
11066 GTGACCATGGCTTGAAGGGCTTGGACGACTCACCAACCCAGTGGGATAACGGCTAC 11007
1072 TTCACCTGCTGTTGGCTACAATTTGGAACTGAAAAGAGTCTCTGCGGTGGCCACCAT 1131
11006 TTTTACTTGTGTTTAAATTTAGCTGGGAATTAAGAGAGTCCCGCGGAGCGGTGGCA 10947
1132 TGGGAACCGATTGACATCAAAAGAAAGAACCGGTTGACGCAAGGACCCCTCTATT 1191
10946 TGGGAACCATCAATATTAAAGAGAAGATTAAGTTATCTCTGTGGGTGACCCCTCGGAG 10887
1192 GGCACAACCCGATCATGACCGATGCGGATATGGCGATAAAGTAAATCCGACCTATCGC 1251
10886 AAGTTTAAACCCCATCATGACGATGCGGATATGCGATGAAGATAGACCCCGGAATCGC 10827
1252 GCTATCTGCGAAAAATTCATGGCGATCTCTGAGTACTTCAAGAAAAAATTTTCGGAAGCG 1311

Db 10826 AATATATCCGAGAGGTTTATATCAAGATCCCGCTATTTCTCCAGAGTGTGTTGCCGTGCT 10767
 QY 1312 TGGTTCAAGCTGACGACCGTGACCTGGGCCCCGAAATACGTTACATCGGCCCGGAAGTG 1371
 Db 10766 TGGTTTAAAGTTAACTCACCGTGATCTTGGGCTTAAAGCGGTTATTTAGGCCAGAGTC 10707
 QY 1372 CCGGCGAAGACCTGATTTGGCAAGACCGGATTCGGCGAG-----GTAACACCGGAC 1422
 Db 10706 CCGAGTGAAGAGTTGATTTGGCAAGACCCCAATCCCGAGCGTCGATTACCGTTAGTTGCC 10647
 QY 1423 TACTGCGAAGAGTGTGCAAGCAAGAAATTTGCACAAAGTGGCTGAGCATTAAGTGAGATG 1482
 Db 10646 AGTGAGATTGTCGAACCTTAGGCGAAATTAATTAGCCTGGGAGCTACGCGTATCGAGTTA 10587
 QY 1483 GTCTCCACCGCTTTGGGACAGTCCCGTACTTATTCGGGGTTCCGATATGCGCGCGGTGCT 1542
 Db 10586 GTGGCGACCGCTTGGGATAGCGCGCGACCTTTTCGAGGCTCGGATTTCCGTGTGGGGCC 10527
 QY 1543 AAGCGTCCCGGATTCGTTTGGCCCCACAGAACGAGTGGCAGGGCAACGAGCGCGGAGCGC 1602
 Db 10526 AATGGCGCACGAATTCGTTTAGCGCCACAAAAGATTGTCAGGAATGAGCCTGAGCGT 10467
 QY 1603 CTGGCGAAGTGTGAGCGTCTACGAGCAGATCTCTCCGACACCGCG-----GCTAGC 1656
 Db 10466 TTGCAAAAGATTGAAGATTATGAGCGGCTTCGAGCGAAGCTTGAGCAAAAAGTCAGT 10407
 QY 1657 ATCGCGAGCGTGTCTTCTGCGCGTAGCGTAGGATCGAGAAAGCGCGAAGCAGCA 1716
 Db 10406 ATTGCTGATTGATTGCTTCGCGCGCGCTGCGCGGATTTGAAAAGCGCGCCACGCTGCG 10347
 QY 1717 GGTACGATGTCGCGCTTCCTTCCTGAAAGCGCGTGGCGATGCGAACCGCGGAGATGACC 1776
 Db 10346 GGAGTCAATATCACAGTACCTTTTCATGCGCTGCGCGTGGTATGCCACGAAGAAATGACC 10287
 QY 1777 GAGCGAGATCCCTTCGACCGCTGGAGCGCTGGCGGATGGCTTCGCGCACTGGCGAGAAG 1836
 Db 10286 GATGTGAATCCCTTGTGCTGTGAACCTTTTCGACGATGCGTATCGCACTGGCTGAAA 10227
 QY 1837 AAAGATATGTGTGAAGCGGGAAGAGATGCTGCTGGATCGTGGCGAGCTGATGGGCTTA 1896
 Db 10226 CAGGATTATGTGTTTACGCTGAGGAGATGATCTCGATCGCTCGCACTGATGGGGTTA 10167
 QY 1897 ACGGGCGGGAATGACGTGCTGCTGGGGGTATCGCGGTACTGGCGACCAACTATGTT 1956
 Db 10166 ACGGCCATGAGATGACGTGATGATGAGCGGATCGGGTATTAGGCACTAACCATGGC 10107
 QY 1957 GGCACCAAAACACGCGCTATTACCGATTGTGAAGCGCAGTTGACCAACGACTTTTGTG 2016
 Db 10106 GGTACGCTACGCTGTGTTTACCGATAAAGTGGGTGTGTTAAGTAATGATTTCCTTTGTG 10047
 QY 2017 AACCTGACCGATATGGGGAACAGCTGGAAGCGCGTAGGTAGCAACCGCTACGAAATCCGC 2076
 Db 10046 AATCTCACCGACATGCGCTACAGCTGGAACCGCTTGGCACTAACCTGTATGAAGTTGT 9987
 QY 2077 GACCGCAAGACCGTCCGTGAAGTGGACCGCTCCGGGTGATCGTGTATTTGTTTC 2136
 Db 9986 GAGCGCAAACTGGTACGTTGAAGTGGAGCGCAACCGCGTCTGATTAGTTGTTGGTCT 9927
 QY 2137 AACTCGCTACTGCGCTTACCGAGAAGTGTAGCGCCGAGGAGATACCGCGGAGAGTTTC 2196
 Db 9926 AACTCTGCTGCGCGCTTACCGCGAGGTGTATGACAGAGGTATGCTAAGGAAAAGTTT 9867
 QY 2197 GTCAGAGACTTCGTCGCCCGCTGGACCAAGTGTAGAACCGCGACCGCTTTCGACGTCGGG 2256
 Db 9866 GTGCATGATTTTGTGCTGCTTGGACCAAGGTGATGAATGCCGACAGGTTTCGATTGCGCT 9807
 QY 2257 TCGTAA 2262
 Db 9806 TAGTCA 9801

RESULT 14
 SYCKATG

LOCUS SYCKATG 2896 bp DNA linear BCT 06-FEB-1999
 DEFINITION Synechocystis sp. catalase hpi (katG) gene, complete cds.
 ACCESSION DB3990
 VERSION DB3990.1 GI:2198405
 KEYWORDS katG; catalase hpi.
 SOURCE Synechocystis sp.
 ORGANISM Synechocystis sp.
 Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 REFERENCE 1 (sites) Ushimaru, T., Hayashi, H. and Murata, N.
 AUTHORS Ushimaru, T., Hayashi, H. and Murata, N.
 TITLE Nucleotide sequence of katG of Synechocystis sp. PCC 6803
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2896) Ushimaru, T.
 TITLE Direct Submission
 JOURNAL Submitted (12-MAR-1996) Takashi Ushimaru, Shizuoka University,
 Department of Biology, Faculty of Science, 836 Oya, Shizuoka,
 Shizuoka 422, Japan (E-mail:sbtushiesci.shizuoka.ac.jp,
 Tel:054-238-4772, Fax:054-238-0986)
 COMMENT On Jun 17, 1997 this sequence version replaced gi:1235737.
 Sequence updated (13-Jun-1997).
 FEATURES
 Location/Qualifiers
 1..2896
 /organism="Synechocystis sp."
 /mol_type="genomic DNA"
 /strain="PCC 6803"
 /db_xref="taxon:1143"
 351..2615
 /gene="katG"
 351..2615
 /gene="katG"
 /codon_start=1
 /transl_table=11
 /product="catalase hpi"
 /protein_id="BA20459.1"
 /db_xref="GI:2198406"
 Translation: "MGTPARKLRNVPFPHNHRKEKPMANDOVPAKCPVMHGANT
 QGNGNLNPNALNDILHQRDKNPMDGDFNYAEAFQQLDLAAVKQDLHLMTDS
 TQSWPADWGHYGLMIRMAHAAQTVRIADRGGAATGNORFAPLNSWPDNLDKAR
 RLWLPFKKYGKNSLWDLILAGTMAYESMLKVVYFAGGREDIHPKEDIYWGAEK
 EWLASDHRYSGESDRESLENPLAAMQGLIYYNPEGVDGHPDPLCTAQDVTTFARMA
 MNDEETVALTAGHTVKGCKNAELIGPEGADVVEGGLGWHNQGKVGRETMS
 SIEGAWTHTPTQWDNGYFMYFLNHEWELKSPAGAWQEPVNIKEEDPVDVDPNI
 RINPIMTDADMAIKDIPIYRISERYEPDYFAEVFAKAKPLHRODLGPKRSYLG
 DVQEDLIWQDPIPVDTLSGEIKELSQILASGLTVSELVCTAWDSARTFRSSDY
 RGANCARLRLPQKNPQNEPTRLKVLAVLENIOANFAKPVSLADLIVLGGGAIA
 KKAALDGLIEVNVVFFLPGRDATQAMDAESFTPLEPHDGYRNWLKQDTAVSPEELLL
 ERTQMLGLTAPEMTVLIGMVRVLGTHNGGTHKGVFTDRVGLNSDFVNLTDMAVQMR
 PAGNNLYEIGDRQTGEVKWTATKVDLVFGSNSILRSYAEVYAQDDNREKFRVDFVAAM
 TKVMNADRFDLPRG"
 BASE COUNT 756 a 724 c 781 g 635 t
 ORIGIN
 Query Match 41.8%; Score 945.2; DB 1; Length 2896;
 Best Local Similarity 65.9%; Pred. No. 1.9e-187;
 Matches 1424; Conservative 0; Mismatches 718; Indels 18; Gaps 3;
 QY 111 CGGTAATCTCTGTCATCGACCGTGGTAAACCTCGACCGGTACTTCCAACAAAGATTG 170
 Db 449 CAGTAATGTCTGTATGACGAGGAGCAATACACCGGCGAGACGGAATCTCAACTG 508
 QY 171 GTGGCGGAGAGGTTTGAACCTGGATATTTTGCATCAGCAAGATCGCAATTCAGACCCGAT 230
 Db 509 GTGGCCCAATGCGCTCAATCTAGACATATTGTCATCAGCATGACCGTAAACAAATCTAT 568
 QY 231 GGATCGGATTTCAACTACCGTGAAGAAAGTAGCAAGTCTGATTTTCGACGCGCTGAAGAA 290
 Db 569 GGACGATGGGTTTAACTATCGCGAAGCGTTTCAACAACTGGATTTAGCTGCTGTTAAACA 628
 QY 291 AGATGTCCACGCGTTGATGACCGATAGCAAGAGTGTGGCCGCTGACCTGGGGGACTA 350
 Db 629 AGACCTACACCATTTGATGACCGATAGCCAAAGATTGTTGGCCCGCGGACTGGGGTCAATTA 688

QY 351 CGGCGTTTGATGATCCGATGCGTGGCACTCCGCTGGCACTACCTACCTACCTATGCTGATGG 410
 Db 689 TGGTGGTAATGATGATGATGCGCTGGCACTCCGCTGGCACTATGCGCATTTGCCGATGG 748
 QY 411 CCGTGGGCGGCTGTGACCGGAAGCAGCGCTTTGACCGCTCAACTCTCGTGGCGGACAA 470
 Db 749 TCGGGGTGGGCGCCACTGGCAACACAGCTTGGCCCCCTGATAGCTGGCCAGATAA 808
 QY 471 CGTCAGCTGGATAAAGCGCGCGCTGCTGTGGCCGATCAAGAAAGAGTACGGCAACAA 530
 Db 809 CGTCAATTTAGACAAGACCGCGCGCTACTCTGGCCGATCAAAAAGAAATATGGCAATAA 868
 QY 531 AATCAGCTGGCAGACCTGATGATCTGCTGGCGACCGCTGCTTATGATGCTCATGCGGCTT 590
 Db 869 ACTTTCTGGGAGATTTGATCATTTTGGCTGGCACCATGGCTTATGATGCTCATGCGGCTT 928
 QY 591 ACCTGCTTACGCTTCTCTTTCGCGCGCTGATATTTGGGAACCCGAAAGAAATATCTA 650
 Db 929 CAAGTTTACGCTTTGCGGCTGGTGGGAAGACATTTGGCACCCGGAAGAGGACATTTA 988
 QY 651 CTGGGGTGACGAAAGAGTGGCTGGCACCTTCTGACGAACGCTACGCGCACTGGAACAA 710
 Db 989 CTGGGGAGCAGAAAGGAATGGCTTCTCCAGTGACCATCGCTACGGCAGTG---AAGA 1045
 QY 711 GCCAGAGACCATGGAACCCGCTGGCGGCTGTCCAAATGGGTCTGATCTATGTAACCC 770
 Db 1046 CCGGGAAGCTTTAGAAAATCCCTTTGGCCCGCTACAAAATGGGATTAATTTACGTTAACCC 1105
 QY 771 GGAAGGTGTTAACGGCCACCTGATCCGCTGAGAACCCGACAGCAGGTACTTGAACCTT 830
 Db 1106 TGAGGGGTGATGGGACCCGGATCTTTATGACGCCACAGGAGCTCCGCACTACCTT 1165
 QY 831 GCCTCGTATGCGGATGACGAGAAACCCGACGCTCTACAGCTGGCGGCGCACACCGT 890
 Db 1166 TGCCCGCATGGCCATGAATGACGAGAAACCGTTGCCCTTACCGCTGGGGGSCACACCGT 1225
 QY 891 CGGTAAATGTCACGTAATGCAATGCCTCTGCTGCTAGCCCTGACCCCAAGGCTCTGA 950
 Db 1226 AGGGAATTGCCATGCAACAGTAAAGCTGAGCTTATAGGGCCAGAACACGAGGGAGCTGA 1285
 QY 951 CGTTGAAACACGAGGCTTAGTTGGGCAACCCCAACATGACGGCAAGCAAGCAACGC 1010
 Db 1286 TGTGTGGAGCAGGCTTGGGTGGCAACCAACCAACGCAAGGTGTGGGCGGGAAC 1345
 QY 1011 CGTGACCTCGGATGCAAGGTGCTTGACACCAACCCACGAAATTCGATATGGGCTA 1070
 Db 1346 TATGTCCAGTGGCATTAAGGGGCTTGGACCAACCCACCTACCCAGTGGGCAACCGTTA 1405
 QY 1071 TTTCGACCTGCTTTCGGCTACATTTGGGMACTGAAAGAGTCTGCGGCTGCCACCA 1130
 Db 1406 TTTTATATGTTGTTCAACCAAGTGGAACTGAAGAAAGTCTCTGCGGGGCTTGGCA 1465
 QY 1131 TTGGGAACCGATTGACATCAAAAGCAACGCGGTGTGACGCGACGACCCCTCTAT 1190
 Db 1466 ATGGGAACCGTCATATTAAGAGAGATTAACCCCTGATGTGGAGGACCCGAATAT 1525
 QY 1191 TGGCCACACCCGATCATGCGGATATGCGGATTAAGGTAAATCCGACCATCG 1250
 Db 1526 TGGCCACCAATCCATCATGACCGATGCGGATATGCCATGATTAAGGACCCCATCTACCG 1585
 QY 1251 CGCTATCTCGGAAATTCATGGCCGATCCTGAGTACTTCAAGAAACTTTCGCGAAGC 1310
 Db 1586 ACAAAATTCGAGCGCTTCTATCGGAGCGCGGATTTATTTGCGGAAGTATTTGCGCAAGC 1645
 QY 1311 GTGGTTCAAGCTGAGCGACCGTGAACCTGGGCGCGGAAATCACCTTACATCGGCGCGGAAGT 1370
 Db 1646 CTGGTTTAACTAACTACAGGATCTAGGACCAAGAGCCGTTACCTTGGCCCCGATGT 1705
 QY 1371 GCCGCGAGAAGACCTGATTTGGCAAGACCCGATTCGCGACGATTAACCCGACTACTGGCA 1430
 Db 1706 GCCCGAGGAAGATTAATTTGGCAAGACCCCATTCGCCAGTGGATTACACCTTTCTGA 1765
 QY 1431 AG-----AAGTGGTCAAGCAGAAAAATTCACAAAATGGCCCTGAGCATTAGTGAGAT 1481

Db 1766 AGGGAAATTAAGGAGCTTGAACAAATAATTTAGCCCTCCGCGCTCACCGTTTCCGAAC 1825
 QY 1482 GGTCTCCACCGCTTGGGACAGTGGCCGCTACTTATCGCGTTCGCGATATCGCGGGGGTGC 1541
 Db 1826 AGTCTGCACTGCTCGGATAGTGGCCGACCTTCGCTAGTCCGACTATCGAGGGGAGC 1885
 QY 1542 TAACGGTCCCGCATTCGCTTGGCCCCACAGAACAGTGGCAGGCAACGAGCCGGAGG 1601
 Db 1886 CAATGGAGCCCGATTCGTTTAAACCCCAAGAAATTTGGCCCGCTAATGAACCAACTCG 1945
 QY 1602 CTTGCGGAAGTCTGAGGCTACGAGCAGATCTCTGC-----CGACACCGGCGCTAG 1655
 Db 1946 ATTGCAAAAGTCTTGGCGGTGCTGAAATAATCAAGCTAATTTTGTAAAGCCGCTAG 2005
 QY 1656 CATCGCGACGTGATCGTTCGCGGTAGCTGAGCATGAGCAAGCGCGGAAAGCAGC 1715
 Db 2006 TCTACGGGACTTAATGTGCTAGGAGGGGAGCGGGATCGCCAAAGCGGCCCTTGATGG 2065
 QY 1716 AGTTTACGATGTCGCGCTTCCCTCTGAAAGCGCTGGCGATGGCAGCCGAGATGAC 1775
 Db 2066 GGGGATCGAGGTTAATGTACTTTTCTGCCGCTCGAGGGATGCCACCAAGCCATGAC 2125
 QY 1776 CGACGACACTCTTCGACCGCTGGAGCGCTGGCGGATGCTTCGCGCAACTGGCAGAA 1835
 Db 2126 CGATCGAATACTTTACTCCGTTGGAACCGATCCACGATGCTTATCGTAACTGGCTAA 2185
 QY 1836 GAAACGATATGTTGGAAGCGGAAGAGATGCTGCTGGAATCGTGGCAGCTCATGGCTT 1895
 Db 2186 ACAGGATTAAGCTGTGTCCCGGAGAAATGCTGTTGGAGCGAACCAATTAATGGGACT 2245
 QY 1896 AACCGCCCGGAAATGACCGTCTGCTGGCGGTATGCGCTACTGCGGACCAACTATGG 1955
 Db 2246 AACGCGACCGAAATGACTGTTTAACTCGTGTATGCGGGTGTGGGTACTAACCATGG 2305
 QY 1956 TGGCACCAACACGCGGTATTCACGATTTGTAAGGCCAGTTGACCAAGCACTTTTGT 2015
 Db 2306 CGGGACTAAGCATGAGTCTTTACCGATCGGGTGGGAGTACTCAGTAACGACTTCTTGT 2365
 QY 2016 GAACCTGACCGATATGGGAACAGCTGGAAGCGGTAGGTAGTACCAAGCGCTACGAAATCG 2075
 Db 2366 TAATTTGACTGATAGGCTACCAATGGCGGCGGCGGCAACATCTCTAGAAATGG 2425
 QY 2076 CGACCGCAAGACCGGTGCGGTGAAGTGGACCGCTCGCGGGTGGATCTGGTATTTGGTTC 2135
 Db 2426 CGATCGCAACACAGGGAAGTGAATGGACTGCCACCAAGTAGACTGCTGTTGGTTC 2485
 QY 2136 CAACCTGCTACTGCGCTCTTACGAGAGTGTACGCCAGGAGCTAACCGCGGAGATTT 2195
 Db 2486 CAACCTCCATCTTACGCTCTTACGAGAGTGTATGCCAGGAGCTAACCTCGGGAATAAT 2545
 QY 2196 CGTCAGAGACTTCGTCGCGCTGGACCAAGTGTGAACCGCGACCGCTTTCGACGCTCG 2255
 Db 2546 TGTCGCTGATTTTGGCGGCTGGACTAAAGTGAATGCGGACCGGTTGATCTCC 2605

RESULT 15

D90910/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

D90910 146174 bp DNA linear BCT 04-JUL-2001
 Synecocystis sp. PCC 6803 DNA, complete genome, section:12/27,
 1430419-1576592.
 D90910 AB001339 BA000022
 D90910.1 GI:1652956
 Synecocystis sp. PCC 6803
 Synecocystis sp. PCC 6803
 Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
 1
 Kaneko,T., Tanaka,A., Sato,S., Kotani,H., Suzuki,T., Miyajima,N.,
 Suglura,M. and Tabata,S.
 Sequence analysis of the genome of the unicellular cyanobacterium
 Synecocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
 region from map positions 64% to 92% of the genome

JOURNAL	DNA Res. 2 (4), 153-166 (1995)		
MEDLINE	96127529		
PUBMED	8590279		
REFERENCE	2		
AUTHORS	Kaneko,T., Sato,S., Kotani,H., Tanaka,A., Asamizu,E., Nakamura,Y., Miyajima,N., Hirosewa,M., Suglura,M., Sasamoto,S., Kimura,T., Hosouchi,T., Matsuno,A., Muraiki,A., Nakazaki,N., Naruo,K., Okumura,S., Shimo,S., Takeuchi,C., Wada,T., Watanabe,A., Yamada,M., Yasuda,M. and Tabata,S.		
TITLE	Sequence analysis of the genome of the unicellular cyanobacterium <i>Synechocystis</i> sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions		
JOURNAL	DNA Res. 3 (3), 109-136 (1996)		
MEDLINE	97061201		
PUBMED	8905231		
REFERENCE	3 (bases 1 to 146174)		
AUTHORS	Tabata,S.		
TITLE	Direct Submission		
JOURNAL	Submitted (28-JUN-1996) Satoshi Tabata, Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research; Yana 1532-3, Kisarazu, Chiba 292-0812, Japan (E-mail: tabata@kazusa.or.jp. URL: http://www.kazusa.or.jp/cyano/, Tel: 81-438-52-3933 (ex. 2330), Fax: 81-438-52-3934)		
COMMENT	Potential protein coding regions were assigned on the basis of similarity search of the ORFs and Genemark analysis.		
FEATURES	Location/Qualifiers		
source	1..146174		
	/organism="Synecocystis sp. PCC 6803"		
	/mol_type="genomic DNA"		
	/strain="PCC6803"		
	/db_xref="taxon:1148"		
gene	/note="synonym: Synechocystis PCC6803"		
	1..486		
	/gene="apca"		
CDS	1..486		
	/gene="apca"		
	/note="ORF_ID:slr2067"		
	/codon_start=1		
	/transl_table=11		
	/product="allophycocyanin a chain"		
	/protein_id="BA017874.1"		
	/db_xref="GI:1652957"		
	/translation="MSIVKTSIVNWADAEARYLSPGELDRKAFVTGGAKRLRAETLT GSRTIVKQAGDRLFQKRPDIIVSPGNAYGEEMTATCLRMDYYLRLVTVGVVSGDVT PTEELGVGVREYRSLGTPEAVALQSVREMKVAVSLGMSDDAAEASAFYDFVICKM S"		
gene	582..1067		
	/gene="apcb"		
CDS	582..1067		
	/gene="apcb"		
	/note="ORF_ID:slr1986"		
	/codon_start=1		
	/transl_table=11		
	/product="allophycocyanin b chain"		
	/protein_id="BA017875.1"		
	/db_xref="GI:1652958"		
	/translation="MQDAITAVINSADVQKYLDAAMDKLSYFASGELRVRAASVI SANATIVKEAVAKSLLYSDVTRPGNNMYYTRRYAACIRDLVYLVATYAMLAGDAS ILDERVLNGLAKETNSLGVPSISVQAIKKEVTASLVGADAGKENGYYLDYICSL S"		
gene	1280..1483		
	/gene="apcc"		
CDS	1280..1483		
	/gene="apcc"		
	/note="ORF_ID:ssr3383"		
	/codon_start=1		
	/transl_table=11		
	/product="phycobilisome LC linker polypeptide"		
	/protein_id="BA017876.1"		
	/db_xref="GI:1652959"		
	/translation="WRMFRITACVPSQTRITQREQLQNTYFTKLVPYDNWFRQQRIW MGGKIVKVEIATGRPGTNAGLA"		
gene	complement(1578..2480)		

gene	/transl_table-11		Db	116042			
------	------------------	--	----	--------	--	--	--

```

Db 114965 ATTGGCAAAAGTCTTGGCGGTGCTGGAATAATATCCAAAGCTAATTTTGCTAAGCCGGTCAG 114906
QY 1656 CATCGCGGACGTATCGTTCTGCGCGGTAGCGTAGGCATCGAGAAAGCCGCGAAAGCAGC 1715
Db 114905 TCTAGCGGACTTAATTTGTGCTAGGAGGGAGCGCGCATCGCCAAAGCGGCCCTTGATGG 114846
QY 1716 AGGTTACGATGTGGCGGTCCCTTCCTGAAAGCGCGGTGCGGATGCGACCGCGCGAGATGAC 1775
Db 114845 GGGGATCGAGGTAAATGATACCTTTTCTGCCCGGTGAGGGGATGCCACCAAGCCATGAC 114786
QY 1776 CGACGCGAGACTCCTTCGCGACCGCTGGACCGGTGGCCGATGGCTCCGCAACTGGCAGAA 1835
Db 114785 CGATGCGAATCCTTTTACCTCGGTGGACCGATCCACGATGGTTATCGTAACGTGGCTAAA 114726
QY 1836 GAAAGAGTATGTGTGGAAGCCGGAAGAGATGCTGTGGATCGTGGCAGCTGATGGCCTT 1895
Db 114725 ACAGGATTACGTGTGTCCCGGAGGAATTGCTGTGGAGCGAACCCTAATTAATGGGACT 114666
QY 1896 AACCGGCCGGAATACCGTGTGCTGGCGGTATGCGCGTACTGGGGACCAACTATGG 1955
Db 114665 AACGGCACCGGAATGACTGTTTAAATCGGTGTATGCGGTGTTGGGTACTAACCATGG 114606
QY 1956 TGGCACCAACACGGCGTATTCACCGATTGTGAGCCAGTTGACCAACGACTTTTTTGT 2015
Db 114605 CGGGACTAAGCATGGAGTCTTTACCGATCGGTGGGAGTACTCAGTAACGACTTCTTTGT 114546
QY 2016 GAACCTGACCGATATGGGGAACAGCTGGAAGCGGTAGTAGCAAGCCCTAGCAAAATCG 2075
Db 114545 TAATTTGACTGATATGGCCTACCAATGGCGGCCAGCGGCAACAATCTCTACGAAATTGG 114486
QY 2076 CGACCGCAAGACGGGTGCGGTGAAAGTGGACCGCCTGCGGGGTGGATCTGTTATTTGGTTC 2135
Db 114485 CGATCCCCAACAGGGGAAGTGAATGGACTGCCACCAAGTAGACCTGGTGTGGTTC 114426
QY 2136 CAACCTGCTACTGCGCTCTTACCGACAAGTGTACGCCAGGACGATAACCGCGGAGATT 2195
Db 114425 CAACCTCATCCCTACGCTCTCTACGCAAGTGTATGCCAGGACGACAAATCGGGGAAAAAT 114366
QY 2196 CGTCAGAGACTTCGTCCGCGCTGGACCAAAAGTGTGAACGCGCGACCGTTTCGACGTCC 2255
Db 114365 TGTCCGTGATTTTGTGGCGCGCTGGACTAAAGTGAATGCGGACCGGTTTGATCTCCC 114306

```

Search completed: October 8, 2003, 12:53:40
Job time : 8212.54 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2003, 06:24:20 ; Search time 581.585 Seconds
(without alignments)
10499.111 Million cell updates/sec

Title: US-09-884-889-5
Perfect score: 2262
Sequence: 1 atgataacgcatcgctga.....gttcgacgctgcgtctgtaa 2262

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2262	100.0	2262	19 AAV06554	Alcaligenes (Deley)
2	2262	100.0	2262	4 ABL46034	DNA encoding Alca
3	1098	48.5	2839	24 ABQ78599	Nucleotide sequenc
4	703	31.1	2196	11 AAQ06815	Sequence encoding
5	672.2	29.7	2238	19 AAV06555	Microscilla furves
6	672.2	29.7	2238	24 ABS53945	DNA encoding Micro
7	620.2	27.4	2331	20 AAX00818	M. tuberculosis ca
8	618.6	27.3	2331	18 AAT90400	Mycobacterium tube

9	618.6	27.3	2331	20 AAX00817	M. tuberculosis ca
10	618.6	27.3	82993	24 ABX09140	Mycobacterium tube
11	618.6	27.3	4403765	22 AAI99683	Mycobacterium tube
12	618.6	27.3	4411529	22 AAI99682	Mycobacterium tube
13	606.6	26.8	2235	19 AAT89967	Mycobacterium tube
14	578.2	25.6	2221	19 AAT89968	Mycobacterium tube
15	578.2	25.6	4795	14 AAQ51531	Mycobacterium tube
16	578.2	25.6	4795	20 AAX16969	M. tuberculosis H37
17	412.6	18.2	1852	21 AAF07490	Fusarium venenatum
18	299.2	13.2	5059	23 AAS89244	DNA encoding novel
19	275.6	12.2	522	24 ABQ78598	Fragment of Agroba
20	270.4	12.0	824	23 AAS73010	DNA encoding novel
21	224.6	9.9	1489	21 AAZ36101	Nucleic acid seque
22	186.4	8.2	671	21 AAF14551	Aspergillus oryzae
23	185.2	8.2	660	21 AAA49864	Mycobacterium tube
24	170.2	7.5	6240	23 AAS90138	DNA encoding novel
25	137.2	6.1	620	17 AAT29130	katG gene fragment
26	137.2	6.1	620	17 AAT29132	katG gene fragment
27	135.6	6.0	620	17 AAT29129	katG gene fragment
28	135.6	6.0	620	17 AAT29131	katG gene fragment
29	128.6	5.7	4653	23 AAS73484	DNA encoding novel
30	128.6	5.7	4653	23 AAS89265	DNA encoding novel
31	128.6	5.7	4653	23 AAS89808	DNA encoding novel
32	128.6	5.7	4653	23 AAS90173	DNA encoding novel
33	98.4	4.4	550	24 ABQ37616	Oligonucleotide fo
34	98.4	4.4	550	24 ABQ37617	Oligonucleotide fo
35	95.8	4.2	150	22 AAF95160	katG gene probe #2
36	91	4.0	391	19 AAV70427	DNA sequence of ka
37	91	4.0	391	24 ABL46037	Mycobacterium tube
38	89.4	4.0	391	19 AAV70424	DNA sequence of ka
39	89.4	4.0	391	19 AAV70425	DNA sequence of ka
40	89.4	4.0	391	24 ABL46034	Mycobacterium tube
41	89.4	4.0	391	24 ABL46035	Mycobacterium tube
42	87.8	3.9	391	19 AAV70426	DNA sequence of ka
43	87.8	3.9	391	24 ABL46036	Mycobacterium tube
44	85.2	3.8	585	24 ABQ23530	Oligonucleotide fo
45	85.2	3.8	585	24 ABQ23531	Oligonucleotide fo

ALIGNMENTS

RESULT 1
AAV06554
ID AAV06554 standard; DNA; 2262 BP.

XX AC AAV06554;

XX DT 03-JUL-1998 (first entry)

XX DE Alcaligenes (Deleya) aquamarinus catalase-64CA2 gene.

XX KW Catalase; epoxidation; hydroxylation; biosensor; paper bleaching;
XX KW pasteurisation; ss.

XX OS Alcaligenes aquamarinus.

XX FH Key Location/Qualifiers

FT CDS 1..2262
FT /tag= a
FT /product= "Catalase-64CA2"

XX PN WO9800526-A1.

XX PD 08-JAN-1998.

XX PF 03-JUL-1997; 97WO-US16513.

XX PR 03-JUL-1996; 96US-0674887.

XX PA (RECO-) RECOMBINANT BIOCATALYSIS INC.

XX XX Adhikary RS, Robertson DE, Sanyal I;

XX		WPI; 1998-086953/08.
DR		p-PSDB; AAW33809.
XX		New bacterial catalases, related nucleic acid vectors and
PT		transformed cells - used as oxidising agents and for detecting or
PT		destroying hydrogen peroxide, e.g. in biosensors
XX		Claim 3; Fig 1; 35pp; English.
XX		The present sequence is of the Alcaligenes aquamarinus catalase-64CA2
CC		gene. Fragments of the gene can be used to identify related sequences.
CC		Catalase-64CA2 may be used to catalyse oxidation reactions such as
CC		epoxidation or hydroxylation. The enzyme can also be used to detect or
CC		destroy hydrogen peroxide, e.g. in connection with glyoxylic acid
CC		production, biosensors, contact lens cleaning, pulp/paper bleaching and
CC		pasteurisation of dairy products. Antibodies raised against
CC		catalase-64CA2 can be used to screen libraries for detection and
CC		purification of cells containing the enzyme.
XX		
SQ		Sequence 2262 BP; 553 A; 612 C; 663 G; 434 T; 0 other;
		Query Match 100.0%; Score 2262; DB 19; Length 2262;
		Best Local Similarity 100.0%; Pred. No. 0;
		Matches 2262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1	ATGAATAACGCATCCGCTGCAGTCTACACAGTAGCTTGCGCAGCAAAGATGCAGAGCATTT 60
Db	1	ATGAATAACGCATCCGCTGCAGTCTACACAGTAGCTTGCGCAGCAAAGATGCAGAGCATTT 60
QY	61	GTTCCCTTGGTATGCCAAGGCATAGAGCAATAAGGGAGAGAGGTATGAGCGGTAAATGT 120
Db	61	GTTCCCTTGGTATGCCAAGGCATAGAGCAATAAGGGAGAGAGGTATGAGCGGTAAATGT 120
QY	121	CCTGTCATGCAGCGTGTAAACCTCCACCGGTACTTCCAACAAGATTGGTGCSCGGAA 180
Db	121	CCTGTCATGCAGCGTGTAAACCTCCACCGGTACTTCCAACAAGATTGGTGCSCGGAA 180
QY	181	GGGTGAACCTGGATATTTTGATCATCAAGATCGCAAATCAGACCAGTGGATCCGGAT 240
Db	181	GGGTGAACCTGGATATTTTGATCATCAAGATCGCAAATCAGACCAGTGGATCCGGAT 240
QY	241	TTCNACTACCGTGAAGAAGTACGCAAGCTCGATTTTCAGCGCGCTGAAGAAGATGTCCAC 300
Db	241	TTCNACTACCGTGAAGAAGTACGCAAGCTCGATTTTCAGCGCGCTGAAGAAGATGTCCAC 300
QY	301	CGGTTGATCACCGATAGCCAAAGAGTGTGCCCCGCTGACTGGGGGCACTACGCGCGTTTG 360
Db	301	CGGTTGATCACCGATAGCCAAAGAGTGTGCCCCGCTGACTGGGGGCACTACGCGCGTTTG 360
QY	361	ATGATCCGTATGGCTTGGCACTCCGCTGGCACCTTACCGTATTCGTGATGGCCGTGGGGGC 420
Db	361	ATGATCCGTATGGCTTGGCACTCCGCTGGCACCTTACCGTATTCGTGATGGCCGTGGGGGC 420
QY	421	GGTGGTACCGBAGCCAGCGCTTTTGACCGCTCAACTCTCGGGCGGACAAACGTACGCCTG 480
Db	421	GGTGGTACCGBAGCCAGCGCTTTTGACCGCTCAACTCTCGGGCGGACAAACGTACGCCTG 480
QY	481	GATTAACCGCGCGCTCTGCTGTGGCCGATCAAGAGAAGTACGGCAACAANAATCAGCTGG 540
Db	481	GATTAACCGCGCGCTCTGCTGTGGCCGATCAAGAGAAGTACGGCAACAANAATCAGCTGG 540
QY	541	GCAGACCTGATGATTCCTGGCTGGCACCGTGGCTTATGAGTCCATGGGCTTACCTGCTTAC 600
Db	541	GCAGACCTGATGATTCCTGGCTGGCACCGTGGCTTATGAGTCCATGGGCTTACCTGCTTAC 600
QY	601	GGCTTCTCTTTCGGCCGCGTCGATATTTGGGAACCCGAAAAAGATATCTACTGGGGTGAC 660
Db	601	GGCTTCTCTTTCGGCCGCGTCGATATTTGGGAACCCGAAAAAGATATCTACTGGGGTGAC 660
QY	661	GA AAAAGAGTGGCTGGCACCTCTCTGACGAACCGTACGGCACGTGGAACAAGCCAGAGACC 720
Db	661	GA AAAAGAGTGGCTGGCACCTCTCTGACGAACCGTACGGCACGTGGAACAAGCCAGAGACC 720

QY 1801 GAGCCGCTGCGCGATGCTTCCGCACTGCGCAGAGAAAGATATGTGGTGAAGCCGGAA 1860
 Db 1801 GAGCCGCTGCGCGATGCTTCCGCACTGCGCAGAGAAAGATATGTGGTGAAGCCGGAA 1860
 QY 1861 GAGATGCTGCTGATGCTGCGAGCTGATGGGCTTAACCGCCCGGAAATGACCGTCTG 1920
 Db 1861 GAGATGCTGCTGATGCTGCGAGCTGATGGGCTTAACCGCCCGGAAATGACCGTCTG 1920
 QY 1921 CTGGCGGTATGCGCGTACTGCGCACCACATATGTTGGTGGCCACCAACACCGCGTATTCACC 1980
 Db 1921 CTGGCGGTATGCGCGTACTGCGCACCACATATGTTGGTGGCCACCAACACCGCGTATTCACC 1980
 QY 1981 GATTGTGAAGCCAGTTGACCAACGACTTTTGTGTAACCTGACCGATATGGGGACAGC 2040
 Db 1981 GATTGTGAAGCCAGTTGACCAACGACTTTTGTGTAACCTGACCGATATGGGGACAGC 2040
 QY 2041 TGGAAAGCCGCTAGTAGCAACGCTACCAATCCGACCGCAGACCGTGGCGTGAAG 2100
 Db 2041 TGGAAAGCCGCTAGTAGCAACGCTACCAATCCGACCGCAGACCGTGGCGTGAAG 2100
 QY 2101 TGGACCGCTCGCGGTGGATCTGGTATTTGGTTCCAACTCGGTACTGCGCTCTTAGCCA 2160
 Db 2101 TGGACCGCTCGCGGTGGATCTGGTATTTGGTTCCAACTCGGTACTGCGCTCTTAGCCA 2160
 QY 2161 GAAGTGTACCGCCAGGACGATACGGCGAGAGTTCGTCAAGACTTTCGTGCGCGCTGG 2220
 Db 2161 GAAGTGTACCGCCAGGACGATACGGCGAGAGTTCGTCAAGACTTTCGTGCGCGCTGG 2220
 QY 2221 ACCAAAGTATGAACCGCGACCGCTTTCGACGTGCGCTGTAA 2262
 Db 2221 ACCAAAGTATGAACCGCGACCGCTTTCGACGTGCGCTGTAA 2262

RESULT 2

ABS53944
 ID ABS53944 standard; DNA; 2262 BP.
 XX ABS53944;
 AC
 DT 22-NOV-2002 (first entry)
 XX
 DE DNA encoding Alcaligenes catalase protein 64CA2.
 XX
 KW Gene; ss; marine bacteria; catalase; hydrogen peroxide; enzyme.
 XX
 OS Alcaligenes aquamarinus.
 XX
 FH Key Location/Qualifiers
 CDS 1..2262
 FT /*tag= a
 FT /product= "Catalase protein"
 XX
 US2002102680-A1.
 XX
 PN 01-AUG-2002.
 XX
 PF 19-JUN-2001; 2001US-0884889.
 XX
 PR 03-JUL-1996; 96US-0674887.
 PR 16-OCT-1997; 97US-0951844.
 PR 05-OCT-1999; 99US-0412347.
 XX
 PA (ROBE/) ROBERTSON D E.
 PA (SANY/) SANYAL I.
 PA (ADHI/) ADHIKARI R.
 XX
 PI Robertson DE, Sanyal I, Adhikari R;
 XX
 WP1; 2002-690613/74.
 DR P-PSDB; ABG33040.
 DR
 XX
 PT New purified catalase polypeptide useful for catalyzing the breakdown
 of hydrogen peroxide, for modifying small molecules, and for generating

PT antibodies which bind to the polypeptide -
 XX Claim 1; Fig 5; 44pp; English.
 CC This invention relates to the DNA and protein sequences of a novel
 purified catalase polypeptide. The protein of the invention or its
 homologue is useful for catalyzing the breakdown of hydrogen peroxide.
 CC The protein sequence of the invention is useful for modifying small
 molecules, by mixing the protein or its fragments with a small molecule
 CC to produce a modified small molecule. The protein of the invention is
 CC also useful for catalysing biochemical reactions, for hydrolysing
 CC glycosidic linkages and for generating antibodies which bind
 CC specifically to the protein. The nucleic acid sequences of the
 CC invention is useful as a probe to determine whether a biological sample,
 CC such as a soil sample, contains an organism having the nucleic acid or
 CC an organism from which the DNA was obtained, or for identifying an
 CC isolating related nucleic acids. An antibody that binds the cellulase
 CC protein of the invention is useful in immunofluorescence chromatography
 CC procedures to isolate or purify the protein or for detection of the
 CC expression in a biological sample. The cellulase protein of the
 CC invention is heat stable, is heat resistant, and is able to
 CC renature and regain activity after exposure to temperatures of from
 CC about 60 to 105 degrees. The present sequence represents the DNA
 CC sequence encoding the Alcaligenes aquamarinus catalase protein of the
 CC invention.
 XX SQ Sequence 2262 BP; 553 A; 612 C; 663 G; 434 T; 0 other;

Query Match 100.0%; Score 2262; DB 24; Length 2262;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGAATAACGATCCGCTGACGATCTACAGTAGTCTTGCAGCAAAAGATGCAGAGCATTT 60
 Db 1 ATGAATAACGATCCGCTGACGATCTACAGTAGTCTTGCAGCAAAAGATGCAGAGCATTT 60
 QY 61 GTTCCCTTGGTATCGCCAAAGCATAGAGCAATAAGGAGAGAGAGTATGAGCGGTAATGT 120
 Db 61 GTTCCCTTGGTATCGCCAAAGCATAGAGCAATAAGGAGAGAGAGTATGAGCGGTAATGT 120
 QY 121 CCTGTCAATGACGGTGTAAACCTCGACCGGTACTTCCAAACAAGATTTGGTGGCCGAA 180
 Db 121 CCTGTCAATGACGGTGTAAACCTCGACCGGTACTTCCAAACAAGATTTGGTGGCCGAA 180
 QY 181 GGGTTGAACCTGGATATTTTGCATCAGCAAGATCGCAATCAGACCGATGGATCCGGAT 240
 Db 181 GGGTTGAACCTGGATATTTTGCATCAGCAAGATCGCAATCAGACCGATGGATCCGGAT 240
 QY 241 TTCAACTACCGTGAAGAAGTACGCAAGCTCGATTTCCAGCGCTGGAAGAAGATGTCAC 300
 Db 241 TTCAACTACCGTGAAGAAGTACGCAAGCTCGATTTCCAGCGCTGGAAGAAGATGTCAC 300
 QY 301 GCGTTGATACCGATAGCAAGAGTGTGGCCCGCTGACTGGGGGCGTGGGCGGTTG 360
 Db 301 GCGTTGATACCGATAGCAAGAGTGTGGCCCGCTGACTGGGGGCGTGGGCGGTTG 360
 QY 361 ATGATCCGATATGGCTTGGCACTCCGCTGGCACTTACCGTATGCTGATGGCGGCGG 420
 Db 361 ATGATCCGATATGGCTTGGCACTCCGCTGGCACTTACCGTATGCTGATGGCGGCGG 420
 QY 421 GGTGGTACCGGAAGCCAGCGCTTTGACCGCTCAACTCCTGGCCGGAACAAGTCAAGCTG 480
 Db 421 GGTGGTACCGGAAGCCAGCGCTTTGACCGCTCAACTCCTGGCCGGAACAAGTCAAGCTG 480
 QY 481 GATAAAGCGCGCGCTGCTGTGCGCCATCAAGAAGAAGTACGCAACAATAACAGCTGG 540
 Db 481 GATAAAGCGCGCGCTGCTGTGCGCCATCAAGAAGAAGTACGCAACAATAACAGCTGG 540
 QY 541 GCAGACCTGATGATTCTGGCTGGCACCCTGCTTATGAGTCCATGGCTTACCTGCTTAC 600
 Db 541 GCAGACCTGATGATTCTGGCTGGCACCCTGCTTATGAGTCCATGGCTTACCTGCTTAC 600
 QY 601 GGCTTCTCTTTTGGCCCGCGTGCATATTTTGGGAACCCGAAAAAAGATATCTACTGGGGTGAC 660

Db 601 GGGTTCTCTTTGGCGCGCTCGATATTTGGGAACCCGAAAGATATCTACTTGGGTGAC 660
 QY 661 GAAAGAGCTGGTGGCACTTCTGACGAACGCTACGGCGAGCTGAAACAAGCCAGAGACC 720
 Db 661 GAAAGAGCTGGTGGCACTTCTGACGAACGCTACGGCGAGCTGAAACAAGCCAGAGACC 720
 QY 721 ATGGAAACCCGCTGGCGCTGTCCAAATGGTCTGATCTATGTGAACCCGGAAGGTGT 780
 Db 721 ATGGAAACCCGCTGGCGCTGTCCAAATGGTCTGATCTATGTGAACCCGGAAGGTGT 780
 QY 781 AACGGCCACCTGATCCGCTGAGAACCCGACAGCAGTACTTGAACCTTCCCGCTATG 840
 Db 781 AACGGCCACCTGATCCGCTGAGAACCCGACAGCAGTACTTGAACCTTCCCGCTATG 840
 QY 841 GCGATGAACGAGGAAACCCAGCCCTCACAGCTGGCGGCAACACCCCTCGGTATTTGT 900
 Db 841 GCGATGAACGAGGAAACCCAGCCCTCACAGCTGGCGGCAACACCCCTCGGTATTTGT 900
 QY 901 CACGGTAATGGCAATGCCCTCTCGGTTAGCCCTGACCCAAAGCCTCTGACGTTGAAAC 960
 Db 901 CACGGTAATGGCAATGCCCTCTCGGTTAGCCCTGACCCAAAGCCTCTGACGTTGAAAC 960
 QY 961 CAGGGCTTAGGTTGGGCAACCCACATGCAAGGCAAGGCAACCCCTGACCTCG 1020
 Db 961 CAGGGCTTAGGTTGGGCAACCCACATGCAAGGCAAGGCAACCCCTGACCTCG 1020
 QY 1021 GGTATCGAAGTGTCTGGACCAACCCACGAAATTCGATATGGGCTATTTTCGACCTG 1080
 Db 1021 GGTATCGAAGTGTCTGGACCAACCCACGAAATTCGATATGGGCTATTTTCGACCTG 1080
 QY 1081 CTGTTGGCTACAATTTGGAACTGAAAGAGTCTCTGCGGTGCGCAACATTTGGGAACG 1140
 Db 1081 CTGTTGGCTACAATTTGGAACTGAAAGAGTCTCTGCGGTGCGCAACATTTGGGAACG 1140
 QY 1141 ATTGACATCAAAAGGAAACAGCGGTTGACGCGCAGCAGCCCTCTATTGGCCACAA 1200
 Db 1141 ATTGACATCAAAAGGAAACAGCGGTTGACGCGCAGCAGCCCTCTATTGGCCACAA 1200
 QY 1201 CCGATCATGACCGATGCGGATATGGCGATAAAGTAAATCCGACCTATCGCCTATCTGC 1260
 Db 1201 CCGATCATGACCGATGCGGATATGGCGATAAAGTAAATCCGACCTATCGCCTATCTGC 1260
 QY 1261 GAAATAATCATGGCGATCTCAGTACTCAAGAAACTTTTCGGAAGCGGTGGTTCAAG 1320
 Db 1261 GAAATAATCATGGCGATCTCAGTACTCAAGAAACTTTTCGGAAGCGGTGGTTCAAG 1320
 QY 1321 CTGACGACCGTACCTGGGCGCGAAATCACGTTACATCGGCGCGGAGTCCGCGAGAA 1380
 Db 1321 CTGACGACCGTACCTGGGCGCGAAATCACGTTACATCGGCGCGGAGTCCGCGAGAA 1380
 QY 1381 GACCTGATTTGGCAAGACCCGATTCGCGAGGTAACCCGACTACTGCGAAGAAGTGGTC 1440
 Db 1381 GACCTGATTTGGCAAGACCCGATTCGCGAGGTAACCCGACTACTGCGAAGAAGTGGTC 1440
 QY 1441 AAGCAGAAATTTGCAAAAGTGGCTGAGCATTTAGTGGTCTCCACCCGCTGGGAC 1500
 Db 1441 AAGCAGAAATTTGCAAAAGTGGCTGAGCATTTAGTGGTCTCCACCCGCTGGGAC 1500
 QY 1501 AGTGGCCGTACTTATCGGCTTCCGATATCGCGCGGTGCTAAAGGTCGCCGATTCGC 1560
 Db 1501 AGTGGCCGTACTTATCGGCTTCCGATATCGCGCGGTGCTAAAGGTCGCCGATTCGC 1560
 QY 1561 TTGGCCCCACAGAGAGTGGCAGGGCAACGAGCGCGGCTGCGGAAAGTGTCTGAGC 1620
 Db 1561 TTGGCCCCACAGAGAGTGGCAGGGCAACGAGCGCGGCTGCGGAAAGTGTCTGAGC 1620
 QY 1621 GTCTACGAGCAGATCTCTGCGACACCCGCGGTAGCATCGCGGAGCTGATCTTCTGCGC 1680
 Db 1621 GTCTACGAGCAGATCTCTGCGACACCCGCGGTAGCATCGCGGAGCTGATCTTCTGCGC 1680
 QY 1681 GGTACGCTAGGATCGAGAAACCCGGAAGCAGGCTTACGATGTCGGGTTCCTCTTC 1740

Db 1681 GGTAGCGTAGGCGATCGAGAAAGCCGCGAAGCAGCAGGTTACGATGTGCGCGTTCCTTC 1740
 QY 1741 CTGAAAGGCGTGGCGATGCGACCGCGAGATGACCGAGCAGCTTCCTTCGACCGCTG 1800
 Db 1741 CTGAAAGGCGTGGCGATGCGACCGCGAGATGACCGAGCAGCTTCCTTCGACCGCTG 1800
 QY 1801 GAGCCGCTGGCGATGGCTTCCGCAACTGCGCAACTGCGCAAGAAAGATGTGTGGTGAAGCCGAA 1860
 Db 1801 GAGCCGCTGGCGATGGCTTCCGCAACTGCGCAACTGCGCAAGAAAGATGTGTGGTGAAGCCGAA 1860
 QY 1861 GAGATGCTGCTGGATGGTGGCGAGCTGATGGGCTTAACCGGCCCGGAAATGACCGTGTG 1920
 Db 1861 GAGATGCTGCTGGATGGTGGCGAGCTGATGGGCTTAACCGGCCCGGAAATGACCGTGTG 1920
 QY 1921 CTGGGCGGTATGCGGCTACTTGGGCAACCACTATGTTGGCACCACCAACACCGCGCTATTTCACC 1980
 Db 1921 CTGGGCGGTATGCGGCTACTTGGGCAACCACTATGTTGGCACCACCAACACCGCGCTATTTCACC 1980
 QY 1981 GATTTGTGAAGGCGAGTTGACCAACGACTTTTGTGTAACCTGACCGATATGGGGAACAGC 2040
 Db 1981 GATTTGTGAAGGCGAGTTGACCAACGACTTTTGTGTAACCTGACCGATATGGGGAACAGC 2040
 QY 2041 TGGAGCCGCTAGGTAGCAACGCTTACGAAATCCGCGACCGCAAGCCGCTGCGTGAAG 2100
 Db 2041 TGGAGCCGCTAGGTAGCAACGCTTACGAAATCCGCGACCGCAAGCCGCTGCGTGAAG 2100
 QY 2101 TGGACCGCTCGCGGTGATCTGTTATTTGGTTCGCACTCGCTACTGCGCTTACGCA 2160
 Db 2101 TGGACCGCTCGCGGTGATCTGTTATTTGGTTCGCACTCGCTACTGCGCTTACGCA 2160
 QY 2161 GAAGTGTACGCCAGGACGATAACCGCGAGAAAGTTGTCAGAGACTTCGTCGCCGCTGG 2220
 Db 2161 GAAGTGTACGCCAGGACGATAACCGCGAGAAAGTTGTCAGAGACTTCGTCGCCGCTGG 2220
 QY 2221 ACCAAGTGTATGACGCCGACCGTTTCGACGTCGCGTCGTAA 2262
 Db 2221 ACCAAGTGTATGACGCCGACCGTTTCGACGTCGCGTCGTAA 2262

RESULT 3
 ABQ78599 standard; DNA; 2839 BP.
 ID ABQ78599 standard; DNA; 2839 BP.
 AC ABQ78599;
 XX
 DT 25-NOV-2002 (first entry)
 XX
 DE Nucleotide sequence of Agrobacterium tumefaciens catalase kata gene.
 KW Catalase; Kata; reactive oxygen species; ROS; ROS-inducible promoter;
 H2O2 level; antioxidant; gene; ss.
 OS Agrobacterium tumefaciens.
 XX
 FH Key Location/Qualifiers
 CDS 308..2479
 FT /*tag= a
 FT /product= "catalase kata"
 XX
 PN WO200263032-A1.
 XX
 PD 15-AUG-2002.
 XX
 PF 05-FEB-2002; 2002WO-SG00018.
 XX
 PR 05-FEB-2001; 2001US-266657P.
 XX
 PA (PANS/) PAN S Q.
 XX
 PI Pan SQ;
 XX
 DR WPI: 2002-643421/69.
 P-PSDB; ABB78212.

XX Determining the ability of a compound to remove a reactive oxygen
PT species (ROS), useful in pharmaceutical and nutraceutical fields and
PT measuring activities of antioxidants, comprises providing a cell having
PT a ROS-inducible promoter
XX
PS Disclosure; Page 56-63; 70pp; English.
XX
CC The present sequence encodes *Agrobacterium tumefaciens* catalase (Kata).
CC Kata has a promoter which is induced by a reactive oxygen species (ROS).
CC The Kata gene can be used in the method of the invention. The
CC specification describes a method for determining the ability of a
CC compound to remove a ROS. The method comprises providing a cell having
CC a ROS-inducible promoter operably linked to a reporter gene that is
CC heterologous to the promoter, and exposing the cell to a compound
CC potentially able to remove the ROS. The method is useful in determining
CC the ability of a compound or gene product to remove a ROS and determining
CC the level of the ROS H2O2 inside a cell, to properly assess and label
CC antioxidant products for use as food supplements, natural products or
CC drugs. The method may also be used in selecting for new or improved ROS
CC scavengers.
XX
SQ Sequence 2839 BP; 580 A; 854 C; 843 G; 562 T; 0 other;
Query Match 48.5%; Score 1098; DB 24; Length 2839;
Best Local Similarity 69.6%; Pred. No. 1.5e-301;
Matches 1488; Conservative 0; Mismatches 650; Indels 0; Gaps 0;
QY 112 GGTAAATGCTCTGTCATGACGAGTGTAACTACCTCGACCGGTACTTCCAAAGAAATGG 171
DB 112 GGTAAATGCTCTGTCATGACGAGTGTAACTACCTCGACCGGTACTTCCAAAGAAATGG 171
QY 332 GGCAAGTGTCCGTCATGCTGAGGAGCAATAGCGGCTCGGCAATCGGTACCAATGG 391
DB 332 GGCAAGTGTCCGTCATGCTGAGGAGCAATAGCGGCTCGGCAATCGGTACCAATGG 391
QY 172 TGGCGGAGAGGTTGAACCTGGATATTTTGCATCAGCAAGATCGCAATCAGACCCGATG 231
DB 172 TGGCGGAGAGGTTGAACCTGGATATTTTGCATCAGCAAGATCGCAATCAGACCCGATG 231
QY 392 TGGCGGAGAGGTTGAACCTGGATATTTTGCATCAGCAAGATCGCAATCAGACCCGATG 451
DB 392 TGGCGGAGAGGTTGAACCTGGATATTTTGCATCAGCAAGATCGCAATCAGACCCGATG 451
QY 232 GATCCGGATTTCAACTACCGTGAAGAAAGTACCAAGCTCGATTTTCAGCGCGTGAAGAA 291
DB 232 GATCCGGATTTCAACTACCGTGAAGAAAGTACCAAGCTCGATTTTCAGCGCGTGAAGAA 291
QY 452 GGCACCTCTTCAACTACCGGAGCGGTGAAGACGCTTGATGTGCAAGCCCTCAAGGCC 511
DB 452 GGCACCTCTTCAACTACCGGAGCGGTGAAGACGCTTGATGTGCAAGCCCTCAAGGCC 511
QY 292 GATGTCCACGCTTGTATGACGATAGCAAGAGTGTGGCGGCTGACTGGGGGCACTAC 351
DB 292 GATGTCCACGCTTGTATGACGATAGCAAGAGTGTGGCGGCTGACTGGGGGCACTAC 351
QY 512 GATGTCCGCGCTTATGACGACAGCCAGGAATGTTGGCGGCGGCTGAGGCGAGTTAT 571
DB 512 GATGTCCGCGCTTATGACGACAGCCAGGAATGTTGGCGGCGGCTGAGGCGAGTTAT 571
QY 352 GCGGCTTTGATGATCCGATGAGCTTGGCACTCCGCTGCGACCTACCTGATGCTGATGCG 411
DB 352 GCGGCTTTGATGATCCGATGAGCTTGGCACTCCGCTGCGACCTACCTGATGCTGATGCG 411
QY 572 GTCCGATGATGGCCGCTTACCTGGCATGCGCGGTTCTATCGTGCACAGCGT 631
DB 572 GTCCGATGATGGCCGCTTACCTGGCATGCGCGGTTCTATCGTGCACAGCGT 631
QY 412 CGTGGGGGCTGTACCGAGCCAGCGCTTGTGACCGCTCACTCCTGCGCGGCAAC 471
DB 412 CGTGGGGGCTGTACCGAGCCAGCGCTTGTGACCGCTCACTCCTGCGCGGCAAC 471
QY 632 CCGCGGCGGCCAATACCGGCAACCGGCTTGTGACCGCTCACTCCTGCGCGGCAAC 691
DB 632 CCGCGGCGGCCAATACCGGCAACCGGCTTGTGACCGCTCACTCCTGCGCGGCAAC 691
QY 472 GTACGCTTGGATTAAGCGCGCGCTGTGTGGCGGATCAAGAAGATACGGCAACAA 531
DB 472 GTACGCTTGGATTAAGCGCGCGCTGTGTGGCGGATCAAGAAGATACGGCAACAA 531
QY 692 GTCAACACGACAGGCGCGCGCTGTGTGGCGGATCAAGAAGATACGGCAACAG 751
DB 692 GTCAACACGACAGGCGCGCGCTGTGTGGCGGATCAAGAAGATACGGCAACAG 751
QY 532 ATCAGCTGGGACAGCTGATGATTTCTGGCTGGCAGCGGTGATGATGATGATGATGATG 591
DB 532 ATCAGCTGGGACAGCTGATGATTTCTGGCTGGCAGCGGTGATGATGATGATGATGATG 591
QY 752 ATTTCTGGGCGACCTTATCGCGCTCGCGCGGATGCGCTATGATGATGATGATGATG 811
DB 752 ATTTCTGGGCGACCTTATCGCGCTCGCGCGGATGCGCTATGATGATGATGATGATG 811
QY 592 CCGTGTACGCTTCTTCTTGGCGCGCTGCTGATATTTGGGAACCCGAAAGATATCTAC 651
DB 592 CCGTGTACGCTTCTTCTTGGCGCGCTGCTGATATTTGGGAACCCGAAAGATATCTAC 651
QY 812 AAGACCTTGGCTTCTGCTTGGCGCGGAGACATCTGGGCGCGGAGGAGGCTGATGATG 871
DB 812 AAGACCTTGGCTTCTGCTTGGCGCGGAGACATCTGGGCGCGGAGGAGGCTGATGATG 871
QY 652 TGGGCTGACGAAAGAGTGGCTGGCACCTTCTGACGAACGCTACGGGACGCTGAACAG 711
DB 652 TGGGCTGACGAAAGAGTGGCTGGCACCTTCTGACGAACGCTACGGGACGCTGAACAG 711
QY 872 TGGGCTGACGAAAGAGTGGCTGGCGCGGAGGAGGCTGATGATGATGATGATGATG 931
DB 872 TGGGCTGACGAAAGAGTGGCTGGCGCGGAGGAGGCTGATGATGATGATGATGATG 931
QY 712 CCAGAGACCATGGAACCCGCTGGCGGCTGTCCAAATGGGTCTGATCTATGATGATG 771
DB 712 CCAGAGACCATGGAACCCGCTGGCGGCTGTCCAAATGGGTCTGATCTATGATGATG 771
QY 932 CCGGAGAGCTGGAACCCGCTGGCGGCTGTCCAAATGGGTCTGATCTATGATGATG 991
DB 932 CCGGAGAGCTGGAACCCGCTGGCGGCTGTCCAAATGGGTCTGATCTATGATGATG 991
QY 772 GAAGTGTTAACGGCCACCTGATCCGCTGAGAAACCGCACAGGCTACTTGAACCTTC 831
DB 772 GAAGTGTTAACGGCCACCTGATCCGCTGAGAAACCGCACAGGCTACTTGAACCTTC 831

DB 992 GAAGTGTCAACGGCAAGTCCGATCCGCTGGCGAGCGCGGCGAGATGCGGAAACCTTT 1051
QY 832 GCCCTATGGGATGAACGACGAAACCCGAGCCCTCACAGCTGGGGGCCACACCGCTC 891
DB 1052 GCCCCCATGGGATGATGACGAGAAACCGTTGCCCTGACGCGCGGCGGCGACATC 1111
QY 892 GGTAAATGCTACGGTAAATGGCAATCCCTGTGCTGTAGCCCTTGACCCAAAGCCCTGTGAC 951
DB 1112 GGCAAGTCCCATGGCAATGGCAGTGTGCCAATCTCAGCCCGGATCCGGAAGCTGCTG 1171
QY 952 GTTGAACACAGGCTTAGTTGGGCAACCCCAACATGACGCGGCAAGCAAGCAACGCC 1011
DB 1172 CGGCAATATCAGGGTCTCGGCTGGATCAATACCAAGGCGCGGCGGATTTGGCGGTGACACC 1231
QY 1012 GTGACCTCGGGTATCGAAGTCTTGGACCAACCAACCCCAAGAAATTCGATATGGCTAT 1071
DB 1232 GTGGTGTGGGATTCGAAGGCGCATGGACACCGCAACCAACCAAGTGGGACAAACGGCTTC 1291
QY 1072 TFCGACCTGCTGCTCGGCTACAAATTTGGGAATCTGAAAAAGAGTCTGCGCGGTGCCACCAT 1131
DB 1292 TFCGACATGCTGTTCAAGCAGCAGTGGACCTTGAGCGCACAGCCCGCGGTCATGCA 1351
QY 1132 TGGGAACCATTTGACATCAAAAGAAACCAAGCCGTTGACGCCAGCCCTCTATT 1191
DB 1352 TGGGCGCGGATTTACCATCGCGAAGAAAGACAGCCCTGTTGATGTCGAGGATCGTGATC 1411
QY 1192 CGCCAAACCCCATGATGACCGATCGGATATGGGATAAAGGTAAATCCGACCTATCGC 1251
DB 1412 CGCACCATCCCATGATGACCGACCGGACATGGCCCTGAGGTGATCGATCCGATACCGC 1471
QY 1252 GCTATCTCGGAAAAATTCATGCGCATCTGAGTACTTCAAGAAAACTTTTCGCGAAGCG 1311
DB 1472 GAGATTTGCTGAAGTTCAGGACGATCAGGACCATTTCTCTGATGCTCTTCGCCCGCC 1531
QY 1312 TGGTTCAAGCTCAGCAGCCGTCACCTGGGCCCGAAATCAGCTTACATTCGCGCGCGGAGTG 1371
DB 1532 TGGTTCAAGCTCAGCAGCCGTCACATGGGGCGAAGTCCCTTACGTCGCGCGCGGATGTT 1591
QY 1372 CCGGAGAGAGCTGATTTGGCAAGACCCGATTCGGGAGGTAAACACCCACTACTCGGAA 1431
DB 1592 CCGGCTGAAGACTGATCTGGCAGGATCCGATCCCGGCGAGGCTCCACGAGCTACGATGTC 1651
QY 1432 GAAGTGTCAAGCAAGAAATTCACAAAGTGGCTGAGCATTTAGTGAGATGTTCTCCACC 1491
DB 1652 GCTGCGCTCAAGGCTAAGATCGCTCGCGCTTCTGTCGCGCATCTGTTTCAAGC 1711
QY 1492 GCTTGGGACAGTCCCGTACTTTATCGCGTTCCGATATCGCGCGGCTGCTTACGCTGCC 1551
DB 1712 GCATGGACAGTGGCGGACCTTCCGTTGTTGGCAAGCGGCGGCGGCGCAATGGCGCG 1771
QY 1552 CGCATTTGCTTGGCCCCACAGAACGATGGCGAGGCAACGAGCGGAGCGGCTTGGCGGAA 1611
DB 1772 CGTATTTGCTCGCACCGCAGAGGATTTGGAAGGCAATGAGCCCGCTCTTTCCCGC 1831
QY 1612 GTGCTGAGGCTTACAGCAGATCTCTGCCACACCGCGGCTAGCATTCGCGGAGCTGATC 1671
DB 1832 GTGCTTTCGCTTTCGAGCCGATTTGCCCGCAACCGGTGCAAGCATCCCGATGATGATC 1891
QY 1672 GTTCTGGCGGTAGCGTAGGCATCGAAGAACCGGAAAGCAGCAGGTTACGATGTCGC 1731
DB 1892 GTTCTGGCTGGCAATTTACGCGTGGAGAGCGGCGGCGGCGGCTGTTTCGATATCGCC 1951
QY 1732 GTTCTTCTTCTTGAAGGCGCTGGCGATGGACCGCGGAGATGACCCAGCAGCACTCTTC 1791
DB 1952 GTGCGCTTCCGCGCGGCTGCTGCTGACGCTTCCGCGGACGAGCAGGATGCCAGACTTT 2011
QY 1792 GCACGCTGGAGCGCTGGCGGATGGCTTCCGCACTGCGAAGAAAGATGATGTTGGTG 1851
DB 2012 GCGCGCTTGGAGCGCTGGCGGATGGTTCCCGCACTGGGTGAAGAAGACTATGTCGTC 2071
QY 1852 AAGCCGGAACAGATGCTGCTGGATCGTGGCAGCTGATGGGCTTACCCGCGCGGAAATG 1911
DB 1911 AAGCCGGAACAGATGCTGCTGGATCGTGGCAGCTGATGGGCTTACCCGCGCGGAAATG 1911

Db	2072	AGCCCCGAAGAGCTGCTGCTGATCGGGGACAGAGCTTCTTGCCCTCACCGGCGCCGGAATC	2131
QY	1912	ACCGTCTGCTGGGGGCTATGGCGCTACTGGGCACCAACTATGGTGGCACCACAAACAGCGC	1971
Db	2132	ACCGTCTCATCGGGGCTGGCGCTCATCGCGCAATACGGCGGTGCGGGCCATGGC	2191
QY	1972	GTATTCACGATTTGTAAGGCGAGTTGACCAACGACTTTTGTGTAACCTGACCGATATG	2031
Db	2192	GTCTTCACCGATAAGCGGGGGCGCTTACAAACGGAGCTTCTTCACGACGTTTGACGGACATG	2251
QY	2032	GGGAACAGCTGGAAGCGGTAGGTAGCAACGCCCTACGAATCCCGGACCCGAACAGCGGT	2091
Db	2252	GCCTATTCCTGGGTCCCGACCGGCAACAATCTCTATGAGATCCGTGATCGCAAGACGGC	2311
QY	2092	GCGGTGAAGTGGACCGCTCGCGGTGGATCTGGTATTTGGTTTCCAACTCGCTACTCGGC	2151
Db	2312	GCAGCCAGATATTCGGCAACCGCGTCGATCTCGTATCGGCTCCAATCCATCCTCGGC	2371
QY	2152	TCATTACGAGAAGTGTACGCCAGGACGATAACGGCGAGAGTTGTCGAGAGACTTCGTC	2211
Db	2372	GCTTATCGGAAGTTTATGGCGAGGACGACCAACAGGGGAAAAATTCGCCCGGACTTCATT	2431
QY	2212	GCGGCTGGACCAAGGTGATGAACGCCGACCGTTTCGA	2249
Db	2432	GCGGCTGGACGAAGGTGATGAACGCCGACCGTTTCGA	2469
RESULT 4			
AAQ06815			
ID	AAQ06815 standard; DNA; 2196 BP.		
XX			
AC	AAQ06815;		
XX			
DT	06-MAR-1991 (first entry)		
XX			
DE	Sequence encoding heat resistant peroxidase.		
XX			
KW	pod10; E.coli UM228; ds.		
XX			
OS	Bacillus stearothermophilus.		
XX			
Key	Location/Qualifiers		
FT	1..2193		
FT	/*tag= a		
XX			
PN	JP02268684-A.		
XX			
PD	02-NOV-1990.		
XX			
PF	07-APR-1989; 89JP-0089469.		
XX			
PR	07-APR-1989; 89JP-0089469.		
XX			
PA	(TOYM) TOYOBO KK.		
XX			
DR	WPI; 1990-372008/50.		
DR	P-PSDB; AAR08205.		
XX			
PT	DNA contg. genetic information of heat resistant peroxidase -		
PT	prepd. by culturing transformant and collecting heat resistant		
PT	peroxidase		
XX			
PS	disclosure; Fig 2; 8pp; Japanese.		
XX			
CC	Sequence may be used to construct plasmid pod10 used to transform		
CC	E.coli UM228. The transformed expression system may be used to		
CC	produce heat resistant peroxidase in large quantities.		
XX			
SQ	Sequence 2196 BP; 576 A; 564 C; 633 G; 423 T; 0 other;		
Query Match			
Best Local Similarity 31.1%; Score 703; DB 11; Length 2196;			
Matches 1326; Conservative 0; Mismatches 765; Indels 44; Gaps 12;			

Qy	138	TAACACCTCGACCGGTACTTCCAAAGAAGATTGGTGGCCGGAAGGGTTGAACCTGGATAT	197	
Db	60	TCAGTCTTCGAATCGAAGAGCAAGAACTGGTGGCCGAACAGCTGAACCTTAAGCAT	119	
Qy	198	TTTGATCAGCAAGATCGCAAAATCAGACCCGATGGATCCGGATTTCACTACCGTGAAGA	257	
Db	120	TCTCCATCAACATGACCGAAAAAGAAATCCTCATGATGAAGATTCAACTATGCTGAGGA	179	
Qy	258	AGTACGCAAGCTCGATTTCGACCGCTGAAGAAGATGTCCACCGTTGATGACCGATAG	317	
Db	180	GTTTCAAAAATAGACTATTGGCGCTCAAGAAGATTTCGCAAACTGATACGGAAG	239	
Qy	318	CCAAGAGTGGTGGCCCTGACTGGGGCCTACGCGGTTTGGATGATCCGTTATGGCTTG	377	
Db	240	CCAAGACTGGTGGCGGCGCATATAGCCATACGGGCCGTTGTTATCCGATGGCGTG	299	
Qy	378	GCATCTCGCTGGCACCTACCGTATGCTGATGCGCGTGGGGGGGGTGGTACCGGAAGCCA	437	
Db	300	GCATTCAGCTGGCACGTAACCGCATCGCCGACGCGCGCGCGCGCTTCGACCGGCACGA	359	
Qy	438	GCGCTTTGCACCGCTCAACTCCCTGGCGGACACGCTCAGCTGGATAAAGCGCGCGCT	497	
Db	360	GCGCTTTGCGCGGTTAAACAGCTCGCGGACAAACGCCAACTTGGATAAAGCGCGCGT-T	418	
Qy	498	GCTGTGGCCGATCAAGAAGAAAGTACGGCAACAAAATCAGCTGGGCACACCTGATTTCT	557	
Db	419	GTATGCGCCGATCAAAAAGAAATACGGGAACAAAATCTCTGGCCGATTTGTCATTTT	478	
Qy	558	GGCTGGCACCGTGGCTTATGAGTCATGGCTTACCTGCTTAGCGGTTCTTTTCGGCGG	617	
Db	479	GGCGCGCAATGTCGCTATTGAATCGATGGTGGAAAAACGATTGGATTGGCGCGCGGC	538	
Qy	618	CGTCGATATTGGGAACCCGCAAAAAGATATCTACTGGGTGACGCAAAAAGATGGCTGGC	677	
Db	539	CGTTGAGCTGGCATCCGGAAGAAGACGTTTATGGGGATCGGAAAAAGATGGCTGCG	598	
Qy	678	ACCTTCTGAGAACCGCTACGCGCACGTGAACAGCCAGAGACCATGGAACACCGCTGGC	737	
Db	599	CTCT-----GAACGCTATTCGCGTGATCGGA-----CCTCGAAAAACCGCTGCG	643	
Qy	738	GGCTGCCAAATGGGTCTGATCTATGTGAACCCGGAAGGTGTTAACGGCCACCGCTGATCC	797	
Db	644	CGC-GTGCAATGGGTAAATCTACGTCAACCCAGAAAGGCGGACGCGAAGCGGATCC	702	
Qy	798	GCTGAGAACCGCACAGCAGGTACTTGAACCTTCGCGCGTATGCGATGAACACGAAAA	857	
Db	703	AAAGCAGGGGGATATCCG----CAGAGACGTTCCGCGCGCATGGGGATGAAGATGAAGA	758	
Qy	858	AACCGCAGCCCTCACAGCTGGCGGCCACACCGTGGTAATGTACCGGTAAATGGCAATGC	917	
Db	759	AACGGTCCGCTTGATCGCGCGGTCATACGTTCCGAAAGCGCACCGCGCGCGCTGCG	818	
Qy	918	CTCTGGCTTAGCCCTGACCCAAAGCCTCTGACGTTTGAACACAGGGCTTAGTTGGGG	977	
Db	819	CAGCGACGTCGGTCCCGAGCGGAAGCGCCCGATTTGAAGCGCAAGGGCTGGATGGAT	878	
Qy	978	CAACCCCAACATCGACGGCAAGCAACGCGCTGACCTCGGGTATCCAAAGTGGCTTG	1037	
Db	879	CAGCTCTTACGAAAGGGAAGGAGCGATCGATCAACAAAGCGGCATTGAAGCGCGCTG	938	
Qy	1038	GACCAACCAACCCACGAAATTCGATATGGCTATTTCGACCTGCTTTCGGCTACAAATG	1097	
Db	939	GACCGCGACCCCAACCCAGTGGGATAGTCTGCTACTTTGACATGCTGTTGGCTATGACTG	998	
Qy	1098	GGAACTGAAAAAGAGTCTCCCGGTGCCACCATTTGGAAACCGATGACATCAAAAAGGA	1157	
Db	999	GTGGCTGACGAAAAAGCCACGCGGGCATGGCAATGGATGGCGTCCGACCGGATGAAA	1058	
Qy	1158	AAACAACCGCGTTGACGCGCAGACCGCTCTATTCCCAACACCGCATGACCGATGC	1217	
Db	1059	AGATTGGCTCCGGATCCCGAGGATCCGTCGAAAAAAGTTCCGACGATGATGACGAC	1118	


```
QY 1218 GGATATCGGATAAAGTAATCGACCTATCGCGCTATCTGCGAAATAATTCATGGCGGA 1277
Db 1119 CGATTGGCTTCGGTTTGGCTTACCGCGAATACGAAATAATCGCGCGGATTCCTATCAAAA 1178
QY 1278 TCCTGAGTACTTCAAGAAACTTTCGCGAAGCGTGTCAAGCTGACGACACCGTGACCT 1337
Db 1179 TCCAGAAGAATTTCCGAGAGCGTTTCCCGGGCATGTTCAAGCTTACCATAGGGATAT 1238
QY 1338 GGGCCCGAAATACGTTTACATCGGCGCGGAAGTCCGCGAGACAGACCTGATTTGGCAAGA 1397
Db 1239 GGGACCGAAACAGAGATATCGGCGCGGAAGTTCGGAAGAAGATTTTCATCTCGCAAGA 1298
QY 1398 CCGATTCCGCGAGTAAACACCGACTACTCGGA- - - - -GAACTGGTCAAGCAGAA 1448
Db 1299 TCCGATTCCAGAGTCGATTACGAATTCAGAGAAGCGGAATTCAGAAATCAAGAAGCAA 1358
QY 1449 AATTGCACAAAGTGGCTGAGCATTTAGTATGATGCTCCACCGCTGGGACAGTGCGCG 1508
Db 1359 AATTTGACACTCGGCGCTGACCGTCACTGACATTTGTGAAACACGTTGGG-CCTCGGCAG 1417
QY 1509 TACTTATCGGGTTCCGATATGCGGCGGTGCTTAACGGTGCCCGCATTCGCTTGGCCCC 1568
Db 1418 CACGTTCCGCAACTCGGATAAGCGCGCG- -ACGAACGCGCGCGCATTCGTCGCGGCC 1475
QY 1569 ACAGAACGAGTGGCAGGCGACGACGCGGACCGCGCTGGCGAAAGTGTGACCGTCT- - - - 1624
Db 1476 GCAAAAAGACTGGGAAGTGAACGACGAGCGGCTCGCCAAAGTGTGCTCGCTTACG 1535
QY 1625 - - -ACGAGCAGATCTCTGCGACACCGCGGTAGCATCGGAGCGTGTGCTTGGCGG 1681
Db 1536 AGGACATCCAGCGGAATCGCGAAAGATGACATTCGCGCATTCGCTTGGCGG 1595
QY 1682 GTACGTTAGGCATCGAGAAAGCCGCGAAGCAGAGTTACGATGTGCGGTTCCCTTCC 1741
Db 1596 CACGCTCGG--TGAAAGGACGCGCGCGCGGCTTTGATGTCAAATTCGCAATTT 1653
QY 1742 TGAAGCGCTGGCGATCGGACCGCGGAGATGACCGAGAGACTCTTCCACCGCTGG 1801
Db 1654 TCCGTGCGCGCGGATGCGACACAGCAAGCAACCGATGTGGAAGCTTTGCGCGTTGG 1713
QY 1802 AGCGCTGGCGGATGGCTTCCGCAACTGGCAGAAAGAGATGTGTGTAAGCGGAAG 1861
Db 1714 AACGTTTCGAGATGGCTTCGCAACTATCAAAAGCAAGATACAGGTTCCGCGGAAG 1773
QY 1862 AGATGCTGCTGGATGCTGCGCAGCTGATGCGGCTTAACCGCGCGGAAATACCGTGTGC 1921
Db 1774 AGCTGCTGCTGACAAAGCCAGCTCTCTCGGC-TGACGCGCCAGAAATGACGCTTAG 1832
QY 1922 TGGCGGTATCGCGTACTGGCACCAACTATGTGGCACCACAAACAGCGGCTATTCACCG 1981
Db 1833 TTGGCGGTTTGGCGGTGTTGGGC-CGAACATATCGGCATCTGCCCTCATGGCGTCTCACTG 1891
QY 1982 ATTGTGAAGCGCAGTTGACCAACACTTTTGTGAACTGACCGATATGGGAACAGCT 2041
Db 1892 ACCGATCGGGGTGCTGACCAACAGCTTCTTTGCACTTGTGGATATGAATATGAAT 1951
QY 2042 GGAAGCCGGTAGGTAGTACCAACGCTTACGAAATCCGCGACCGCAAGACCGGTCCGTTGA 2101
Db 1952 GGGTCCGACAGACAGCGGCTTTATGAATCCGCGCGGAAACGCGGCGGAGTGGGT 2011
QY 2102 GGACGCGCTCGCGGTGATGTTGTTTGGTTTCCAACTGCTGCTACTGCGCTTCTACGCGAG 2161
Db 2012 GGACGCGACCGGGTGGATCTCATTTTCGGATTTTACGCCCAAGACGACCAACCAAGAG 2071
QY 2162 AAGTGTAGCCCGACGACATACGCGGAGAGTTGCTCAGAGACTTCTGTCGCGCGCTGGA 2221
Db 2072 AATTTTACGCCCAAGACGACCAACCAAGAAATTTGTCGCTGATTTTCATCAACGCTGGG 2131
QY 2222 CCAAAGTATGAACGCCGACCGTTTTCGACGTCGCG 2256
Db 2132 TGAAGTATGATGACCGGACCGCTTTGATCTGGTG 2166
```

```
RESULT 5
AAV06555
ID AAV06555 standard; DNA; 2238 BP.
XX AC AAV06555;
XX DT 03-JUL-1998 (first entry)
XX Microscilla furvescens catalase-53cal gene.
XX Catalase: epoxidation; hydroxylation; biosensor; paper bleaching;
XX pasteurisation; ss.
XX Microscilla furvescens.
XX OS
XX Key Location/Qualifiers
XX CDS 1..2238
XX FT /*tag= a
XX FT /product= "Catalase-53cal"
XX PN WO9800526-A1.
XX PD 08-JAN-1998.
XX PF 03-JUL-1997; 97WO-US16513.
XX PR 03-JUL-1996; 96US-0674887.
XX PA (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX PI Adhikary RS, Robertson DE, Sanyal I;
XX WPI; 1998-086953/08.
XX P-PSDB; AAW33810.
XX PT New bacterial catalases, related nucleic acid vectors and
XX transformed cells - used as oxidising agents and for detecting or
XX destroying hydrogen peroxide, e.g. in biosensors
XX Claim 3; Fig 2; 35pp; English.
XX The present sequence is of the Microscilla furvescens catalase-53cal
XX gene. Fragments of the gene can be used to identify related sequences.
XX Catalase-53cal may be used to catalyse oxidation reactions such as
XX epoxidation or hydroxylation. The enzyme can also be used to detect or
XX destroy hydrogen peroxide, e.g. in connection with glyoxylic acid
XX production, biosensors, contact lens cleaning, pulp/paper bleaching and
XX pasteurisation of dairy products. Antibodies raised against
XX catalase-53cal can be used to screen libraries for detection and
XX purification of cells containing the enzyme.
XX SQ Sequence 2238 BP; 634 A; 545 C; 605 G; 454 T; 0 other;
Query Match 29.7%; Score 672.2; DB 19; Length 2238;
Best Local Similarity 59.5%; Pred. No. 1.9e-180;
Matches 1277; Conservative 0; Mismatches 808; Indels 60; Gaps 6;
QY 150 CGGTACTTCCAAACAAGATTGGTGGCGGAAGGGTTGAACCTGGATATTTTCATCAGCA 209
Db 99 CGGCACCAAAACAGGGATTGGTGGCCCAACATGCTCAACCTCGGCATCTTACGCCAACA 158
QY 210 AGATCGCAAAATCAGACCCGATGGATCCGGATTTCAACTACCGTGAAGAAGTACGCAAGCT 269
Db 159 TTCATCGTATCGGACCCCAACGACCCGCGATTTTGACTATGCCGAAGATTTAAGAAGCT 218
QY 270 CGATTTTCAGCGGTGAAGAAGATGTCCACGCTTGTATGACCGATAGCCAGAGATGGTG 329
Db 219 AGATCGCAGCGGTAAAGAGGACCTGGCAGCGCTAATGACAGATTACAGAGACTGGTG 278
QY 330 GCCCGCTACTGGGGGCACTACGGCGGTTTGATGATCGGTATGGCTTGGCACTCCGCTGG 389
Db 279 GCCACGATTACGGTTCATTATGGCCCTTCTTTATACGCATGGCGTGGCAGCGCGG 338
```

QY	390	CACCTACCGTATTCGTGATGCCGTGGGGCGGTGGTACCGGAAGCAGAGCGCTTTGCAAC	449
DB	339	CACCTACCGTATTCGGTGTATGCCGTGGTGGCGGTCCGCGCTACAGCGCTTCGCGCC	398
QY	450	GCTCAACTCCCTGGCGGACACGCTCAGCCTGGATTAAAGCGCGCGCTGTGCTGTGGCGCAT	509
DB	399	TCTCAATAGCTGGCGCAGACAATGCCAATCTGGATTAAAGCAGCGTTGCTTCTTTGGCCCAT	458
QY	510	CAAGAAGAAGTACGGCAACAATAATCAGCTGGCGAGAGCTGATGATTCTTGGCTGGCACCGT	569
DB	459	CAACAAAAATACGGTCCAAAAATCTCTGGCGGATCTAATGATATCTACAGGAACAGT	518
QY	570	GGCTTATGAGTCCATGGGCTTACCTGCTTACGGCTTCTCTTTGGCGCGCTCGATATTG	629
DB	519	AGCTCTGAAACTATGGGCTTTAAAACTTTTGGTTTTGCAGTGGCAGACGATGATG	578
QY	630	GGAAACCGAAAAAGATATCTACTGGGTGACGAAAAAGAGTGGCTGGCACCTTCTGACGA	689
DB	579	GGAGCTTGAAGAGATGTATCTGGGAGCAGAAACCGAATGGCTGGGA-----GACAA	633
QY	690	ACGCTACGGCCACGTGAACAAAGCAGAGACCATGGAAACCGCTGCGGCTGTCCAAAT	749
DB	633	CGGCTATGAAGGTACCCGAGAC-----TCGAAATCCCTGGAGCGCTTACAAT	683
QY	750	GGGTCTGATCTATGAAACCGGAAGTGTTAACGGCCACCTCTGATCCGTGAGAACCGC	809
DB	684	GGGACTCATCTATGTAACACCGGAAGCAACCGCAAGCAGCACACCTATCGCTGCTGC	743
QY	810	ACAGCAGGTACTTTGAACCTTCGCGCGTATGGCCATGAACGACCAAAAAACCGCAGCCT	869
DB	744	CGGTGATTTCGTGAGACTTTTGCCGGAATGGCAATGAATGACGAGAAACCGTGGCTCT	803
QY	870	CACAGCTGGCGGCCACACCGTCGGTAAATGTACGGTAAATGGCAATGCCCTCTGGCT--T	926
DB	804	CATAGCGGGTGACACACCTTCGGAAAAACCCATGGTGTCTGCCGATCGGAGAAATATGT	863
QY	927	AGCCCTTGACCCAAAGAGCTCTGAGTTGAAACACAGGCTTAGTTGGGGCAACCCCAA	986
DB	864	GGGCGGAGAGCTTCGCGCGCAGGTATTGAAGAAATGAGCTGGGGTGGAAAAACACTA	923
QY	987	CATCAGGGCAAGCAAGCAACCGCTGACCTCGGTTATCGAAGTCTTTGGACCACCA	1046
DB	924	CGGCACCGGACCGGTGGGATACCATCACGATGGACTAGAGGGCGCTTGGACCAAGC	983
QY	1047	CCCCACGAAATTCGATATGGGCTATTTCGACCTGCTGTTCGGCTACAAATTTGGGAATCGAA	1106
DB	984	CCCTACTCAATGGAGCAATAACTTTTGAANAACCTCTTGGTTACAGTGGGAGCTTAC	1043
QY	1107	AAAGAGTCTCGCGGTGCCACCAATTTGGGAACCGATTGACATCAAAAGGAAACAAAGCC	1166
DB	1044	CAAAAGTCCAGCTGGAGCTTATCAGTGGAAACCAAAAGACGGTCCCGGGCTGGCACCAT	1103
QY	1167	GGTTGAGCCAGCACCCCTCTATTCCCCACAACCCGATCATGACCGATCGGATATGCC	1226
DB	1104	ACCGGATGCACATGATCCCAAGCAAGTCGACGCTCCATTATTATGCTACTACGAGCCTTGC	1163
QY	1227	GATAAAGTAAATCCGACCTATCCGCTATCTTCGGAATAATTCATGSCCGATCCTGAGTA	1286
DB	1164	GCTCGGCTAGGACCTGATTACGAAAAAATTTCTCGACGGTACTATTGAAAACCCCTGATGA	1223
QY	1287	CTTCAAGAAACTTTTCGGAAGCGTGGTTCAAGCTGACCGACCGTGGCGCCGCGAA	1346
DB	1224	TTTTTGCAGATGCTTTTCGGAAGCATGTTACAAACTGACACACAGAGATATGGACCAA	1283
QY	1347	ATCAGTTTACATCGGCCCGGAAGTGGCGGCAGAAAGCTGATTTTGGCAAGCCGATTC	1406
DB	1284	GGTCGGCTACTGGGACCAAGTGGCTTCAGGAGAGCTCATCTGGCAAGACCTATACC	1343
QY	1407	GGCAGGTTAAC-----ACCGACTACTTCGGAAGAAGTGGTCAAGCAGAAAAATTCG	1454
DB	1344	AGATGTAAAGCATCCTCTGTAGACGAAAAACGATATTGAAGGCCCTAAAGCCAAAAATCCT	1403
QY	1455	ACAAAGTGGCTGAGCATTAGTGAGATGGTCTCCACCGCTTGGGACAGTGGCGGTACTTA	1514

Db	1404	GAATTCGGAGCTGACGGTAAAGCGAGCTGGTAAAGCAGCGCATGGCGCTTCTGCATCTACTTT	1463
Qy	1515	TCGCGGTTCCGATATGCGCGCGGTGTAAACGGTCCCGCATTCGCTTGGCCCCACAGAA	1574
Db	1464	TAGAACTCTGCAAGCGCGCGGTGCCACGGTGCAGGTATACGACTGGCCCCACAAA	1523
Qy	1575	CGAGTGGCAGGCCAACGAGCGCGAGCGCCTGGCGAAAGTGCTGAG-----	1619
Db	1524	AGACTGGGAAGCTAAACAAACCCTCAGCAACTTGCACAGGTACTCTAAACACTAGAAAGCTAT	1583
Qy	1620	-----CGTCTACGAGCAGATCTCTGCGGACACCGCGCTAGCATATCGGGAGACTGAT	1670
Db	1584	CCAGGAGGACTTTAAACGAGCGCAATCAGATAACAAAGCAGTATCGTTGGCCGACCTGAT	1643
Qy	1671	CGTTCTGCGCGGTAGCGTAGCATCGAGAAAGCGCGAAAGCAGCAGGTACGATGTCG	1730
Db	1644	TGTGCTGCCCCGCTGTGCGGCTGTAGAAAAGCTGCAAAAGATGCTGGCCATGAGGTGCA	1703
Qy	1731	CGTTCCCTTCCTGAAAGCCGTGCGGATGGACCGCGAGATACCCAGCAGCAGACTCCTT	1790
Db	1704	GGTGCCYTTCAACCCGGGACGAGCGGATGCCACGCTGAGCAAAACCAGTGTGNAAGCTTT	1763
Qy	1791	CGCACCGCTGGAGCCGCTGGCCGATGCTCCGCACTGCGAGAAAGAGTATGTGGT	1850
Db	1764	CGAAGCACTAGAGCCAGCGGTGACGGCTTAGAAAACTACATTAACCCGGAGCATAAAGT	1823
Qy	1851	GAAGCCGGAAGAGATGCTGTGGATCTGGCAGCTGATGGCTTAACCGGCCGCGGAAT	1910
Db	1824	ATCCGCTGAGGAAATGCTGCTAGACCGGGGCGAGCTTCTGTGCGTTTCGCGACCCAGAAAT	1883
Qy	1911	GACCGTGTGCTGGCGCGTATGCCGCTACTGGGCACCAACTATGGTGGCACCAACACGG	1970
Db	1884	GACTGCTTTGTTAGCGGATGCTGCTGTACTGGGCACCACTACGACGGTTTCGACGATGG	1943
Qy	1971	CGTATTCACCGATTGTGAAGCCAGTTGACCAAGCACTTTTTTGTGAACCTGACCCGATAT	2030
Db	1944	AGTGTTTACAAATAAGCCGGGTACGCTATCCAATGACTTCTTTGTAAACCTGTAGACCT	2003
Qy	2031	GGGGAACAGCTGG-----AAGCCGTTAGGTAGCAACGCCCTACGAAATCCGGACCCGAA	2084
Db	2004	CAACACTAAATGGCGAGCCACGCGATGAATCAGACAAAGTTTGTGAAGCGCAGAGACTTCAA	2063
Qy	2085	GACCGGTGCCGTGAAGTGGAGCCGCTCGCGGGTGGATCTGTTATTTGGTTCCAACTCGCT	2144
Db	2064	AACCTGGCGAAGTAAGTTGGAGTGGCACC CGGGTAGACCTGATCTTCGGATCCAAATCCGA	2123
Qy	2145	ACTCGGCTCTTACGAGAAGTGTACGCCCGAGCAGATACGCGGAGAGTTCGTCTAGAGA	2204
Db	2124	GCTAAGAGCCCCTCGAGAAAGTGTACGGCTGTGCAGATTTCTGAAGAAAGTTTGTGTTAAAGA	2183
Qy	2205	CTTCGTGCGCGCCTGGACCAAGTGTGAACCGCCGACCTGTTTGA 2249	
Db	2184	TTTTGTGAAGGCCCTGGCCCAAGTAATGGACCTGGACCGGTTTGA 2228	

RESULT 6

RESULTS 0
ABS53945
ID ABS53945 standard; DNA; 2238 BP.

AA
AC
ABS53945;

22-NOV-2002 (first entry)

XX DE DNA encoding Microscilla catalase protein 53CA1.

Gene; ss; marine bacteria; catalase; hydrogen peroxide; enzyme.

AA
OS Microscilla furvescens.

XX	Key	Location/Qualifiers
FH		

FT	key	LOCATION/
FT	CDS	1..2238

```

FI      CDS
FT      1..2230
        /*tag= a

```

FT XX /product= "Catalase protein"

PN US2002102680-A1.

XX 01-AUG-2002.

XX 19-JUN-2001; 2001US-0884889.

XX 03-JUL-1996; 96US-0674897.

PR 16-OCT-1997; 97US-0951844.

PR 05-OCT-1999; 99US-0412347.

XX (ROBE/) ROBERTSON D E.

PA (SANY/) SANYAL I.

PA (ADHI/) ADHIKARI R.

XX Roberton DE, Sanyal I, Adhikari R;

XX WPI: 2002-690613/74.

DR P-PSDB: ABG33041.

XX

PT New purified catalase polypeptide useful for catalyzing the breakdown

PT of hydrogen peroxide, for modifying small molecules, and for generating

PT antibodies which bind to the polypeptide -

XX Claim 1; Fig 6; 44pp; English.

PS

XX This invention relates to the DNA and protein sequences of a novel

CC purified catalase polypeptide. The protein of the invention or its

CC homologue is useful for catalysing the breakdown of hydrogen peroxide.

CC The protein sequence of the invention is useful for modifying small

CC molecules, by mixing the protein or its fragments with a small molecule

CC to produce a modified small molecule. The protein of the invention is

CC also useful for catalysing biochemical reactions, for hydrolysing

CC glycosidic linkages and for generating antibodies which bind

CC specifically to the protein. The nucleic acid sequences of the

CC invention is useful as a probe to determine whether a biological sample,

CC such as a soil sample, contains an organism having the nucleic acid or

CC an organism from which the DNA was obtained, or for identifying an

CC isolating related nucleic acids. An antibody that binds the cellulase

CC protein of the invention is useful in immunoaffinity chromatography

CC procedures to isolate or purify the protein or for detection of protein

CC expression in a biological sample. The cellulase protein of the

CC invention is heat stable, is heat resistant, and is able to

CC renature and regain activity after exposure to temperatures of from

CC about 60 to 105 degrees. The present sequence represents the DNA

CC sequence encoding the Microscilla furvescens catalase protein of the

XX invention.

XX

XX Sequence 2238 BP; 634 A; 545 C; 605 G; 454 T; 0 other;

Query Match 29.7%; Score 672.2; DB 24; Length 2238;

Best Local Similarity 59.5%; Pred. No. 1.9e-180;

Matches 1277; Conservative 0; Mismatches 808; Indels 60; Gaps 6;

QY 150 CGGTACTTCCACAAAGATTGGTGGCGGAAGGGTTGAACCTGGATTTTGCATGACGA 209

DB 99 CGGCACCAAAACAGGATTGGTGGCCCAACATGCTCAACCTCGGCATCTTACGCCAACA 158

QY 210 AGATCGCAATACAGACCGGATGGATCGGATTTCAACTACCGTGAAGAAGTACGCAAGCT 269

DB 159 TTATCCCTATCGGACCCCAACGACCCGGATTTGACTATGCGGAAAGTTTAAGAAGCT 218

QY 270 CGATTTTCAGCGGTGAAGAAGATGTCACGCGTTGATACCGCATAGCCAAAGAGTGGTG 329

DB 219 AGATCTGGCAGCGGTTAAAAAGGACCTGGCAGCGCTAATGACAGATTCACAGGACTGGT 278

QY 330 GCGCGTGACTGGGGGCACTACGCGGTTTGATGATCCGTATGCGTGGCACTCCGCTGG 389

DB 279 GCCAGCATATTACGGTCAATTATGGCCCTTCTTTATACGATGCGGTGGCAGCGCGG 338

QY 390 CACCTACCGTATGCTGATGCGCTGGGCGGTGGTACCGGAAGCAGCGCTTTGCACC 449

XX (|||||) (|||||) (|||||) (|||||) (|||||) (|||||) (|||||) (|||||) (|||||) (|||||)

Db 339 CACTACCGGTATCGGTGATGCCCGTGGTGGGGTGGCTCCGGGTCAACAGCGCTTCGGCGCC 398

QY 450 GCTCAACTCTCGCGGACAAACGTCAGCCTGGATAAAGCGCGCGCTCTGCTGTGGCGGAT 509

Db 399 TCTCAATAGCTGGCCAGACAAATGCCAATCTGGATAAAGCAGCGCTTCTTTGGGCCAT 458

QY 510 CAAGAAGAAGTACGGCAACAAATACAGCTGGGAGACCTGATGATTTCTGGCTGGCACCGT 569

Db 459 CAACAACAAATACGGTCAAAAATCTCTGGCGGATCTAATGATACTACAGCAAGAAAGCT 518

QY 570 GGCTTATGAGTCCATGGCTTACCTGTACGGCTTCTCTTTCGGCGCGCTGCATATTG 629

Db 519 AGCTCTGGAACATATGGGCTTTAAACTTTTGGTTTTCAGGTGGCAGACGATGATG 578

QY 630 GGAACCCGAAAAAGATATCTACTGGGTGACGAAAAAGAGTGGTGGCACCTCTTGACGA 689

Db 579 GGAGCCTGAAGAAGATCTATCTGGGAGCAGAACCCGAATGGCTGGGA-----GACAA 632

QY 690 ACCTACGGCGAGTGAACAAAGCCAGAGACCATGGAACCCGCTGGGGGTGTCCAAAT 749

Db 633 CGCGTATGAAGGTGACCGGAGC-----TCGAAATCCCTGGGAGCGGTACAAAT 683

QY 750 GGGTCTCATCTATGTGAACCGGAAGGTGTTAAGCGCCACCTCATCGCTGAGAACGC 809

Db 684 GGGACTCATCTATGTAACCCCGAAGGACCCCAAGCAGCAGCCATATCGCTGTGTC 743

QY 810 ACAGCAGGTACTTGAACCTTCGCCCTGATGGCGATGAACGACGAAAAACCCGAGCGCT 869

Db 744 CGCTGATATCGTGAGACTTTGGCCGAATGGCAATGAATGACGAAGAACCGTGGCTCT 803

QY 870 CACAGTGGCGGCGCACACCGTCTGGTAATGTGACGGTAATGGCAATGCTTGGCT--T 926

Db 804 CATAGCGGTGGACACACCTTCGGAAAAACCCATGGTGTCTGCCGATCGGAGAAATATGT 863

QY 927 AGCCCTGACCCAAAGCCTCTGAGTGTGAACACAGGCTTAGTGTGGGCAACCCCAA 986

Db 864 GGGCGGAGAGCCTCGCCGCGAGGTATTTGAAGAAATGAGCCTGGGTGGAAAAACACCTA 923

QY 987 CATGACGGGCAAGGCAAGCAACGCGTGACCTCGGGTATCGAAGTGTCTTGGACACCAA 1046

Db 924 CGGCACGGACACGCTGGGATACCATCACCAAGTGGACTAGAAGCGCTGGACCAAGAC 983

QY 1047 CCCACAAATTCATATGGCTATTTGACGCTGCTGTGGCTGACAAATGGGAACGTAA 1106

Db 984 CCCTACTCAATGGAGCAATAACTTTTGAAGACCTCTTTGGTTACGAGTGGGAGCTTAC 1043

QY 1107 AAAGAGTCTCGCGGTGCCACCATTTGGAAACCCATTGACATCAAAAGAAAGAAACAGCC 1166

Db 1044 CAAAGTCCAGCTGGAGCTTATCAGTGGAAACCAAAAGACCGGTGCGGGGTGGCACCAT 1103

QY 1167 GGTGACGCCAGCACCCCTCTATTTCGCCCAACCCGATCATGACCGATGCGGATATGGC 1226

Db 1104 ACCGATGCACATCATCCAGCAAGTCCGACGCTCCATTTATGCTCACTACGAGACCTGGC 1163

QY 1227 GATAAAGGTAAATCCGACCTATCGGCTATCTGCGAAAAATTTCAATGCGCGATCTGAGTA 1286

Db 1164 GCTCGCATGGACCTGATTTACGAAAAATTTCTCGACGCTACTATGAAAAACCCCTGATGA 1223

QY 1287 CTTTCAAGAAAAATTTTCGGAAGGGTGGTTCAAGCTGACGCAACCGTGAACCTGGGCCGAA 1346

Db 1224 GTTTGCAAGTCTTTCCGGAAGCATGTGTACAACTGACACACAGAGATATGGGACCAAA 1283

QY 1347 ATCAGTTACATCGGCCCGGAAGTGGCGGAGAGACCTGATTTGGCAAGACCCGATTC 1406

Db 1284 GGTGCGCTACCTGGGACCAAGTGGCTTCAGGAAGACCTCATCTGCGAAGACCCCTATACC 1343

QY 1407 GCGAGGTAAAC-----ACCGACTACTGCGAAGAGTGTGTCAAGCAAGAAATTCG 1454

Db 1344 AGATGTAAGCCATCCTCTTGTAGACGAAACGATATTGAAGGCCCTAAAAAGCAAAATCCT 1403

QY 1455 ACAAGTGGCTGAGCATTTAGTGAGATGGTCTCCACCGCTTGGGACAGTGCCTTACTTA 1514

Db 1404 GGAATCGGGACTGACGGTAAGCGAGCTGGTAAGCAGCGCATGGGCTTCTGTCATCTACTTT 1463

QY 460 TGGCGGACAAAGCTCAGCCTGGATAAAGCGCGCGCTGTGCTGTGGCCGATCAAGAAG 519
 Db 472 TGGCGGACAAAGCTCAGCCTGGATAAAGCGCGCGCTGTGCTGTGGCCGATCAAGAAG 531
 QY 520 TACGGCAAAATACAGCTGGCAGACCTGATGATCTGGCTGGCAGCAGCTGATGAG 579
 Db 532 TACGGCAAAAGCTCAGCTGGCAGACCTGATGATCTGGCTGGCAGCAGCTGCGCTGGA 591
 QY 590 TCCATGGGCTACCTGCTTACGGCTTCTTTTCGGCGCGCTGATATTTGGGAACCCGAA 639
 Db 592 TCGATGGGCTTCAAGACCTTCGGGTTTCGGCTTCGGCGCGGTCGACAGTGGGCGCG 649
 QY 640 AAGATATCTACTGGGCTGAGAAAGAGTGGCTGGCAGCCTCTGACGAAAGCTACGGC 699
 Db 650 -ATGAGGCTATATGGGCAAGAACCCACCTGGGTCG-----GCGATGAGCGTTACAGC 702
 QY 700 GACGTGAACAGCAGACCATGAAACCCCGTGGCGGCTGTCCAAATGGTCTGATC 759
 Db 703 GGTAAAGCGGATC-----TGGAGAACCCGCTGGCGCGGTCAGATGGGCTGATC 753
 QY 760 TATGTGAACCCGGAAGTGTAAAGGCGCACCTGATCCGCTGAGAACCGCAGAGGTA 819
 Db 754 TACGTGAACCCGGAAGGCGCAAGCGCAACCGGACCCATGCGCGCGGCTGCACAT 813
 QY 820 CTTGAACCTTCGCGCTATGGGATGAACGACGAAACCGGAGCCCTCACAGCTGGC 879
 Db 814 GCGAGAGCTTCGCGCATGGCCATGAACGACCTCGAAACAGCGGCGCTGATCGTCGGC 873
 QY 880 GCCACACCGTCGGTAATGTGACGGTAATGGCAATGCTCTGCTTAGCCCTGACCCA 939
 Db 874 GGTACACCTTCGGTAAGACCATGGCGCGCGCGCGGATCTGGTCGGCGCCGAAACC 933
 QY 940 AAGGCTCTGAGTTGAAACCAAGGCTTAGGTTGGGCAACCCCAACATCGAGGGCAAG 999
 Db 934 GAGGCTGCTCGGCTGGAGCAGATGGCTTGGCTGGAAGAGCTCGTATGGCACCGGAACC 993
 QY 1000 GCAAGCAACCGCTGACCTCGGATTCGAAGTCTCTGGACCAACCAACCCACGAAATTC 1059
 Db 994 GGTAGAGCGGATCACCACCGGCATCGAGGTCGTATGGAGCAACCCCGCAAAATGG 1053
 QY 1060 GATATGGGCTATTCGACCTGCTGTTCGGCTACAAATGGGAATGAAAGAGTCTCGCC 1119
 Db 1054 GACAAAGTTTCCTCGAGATCTGTACGCTACGAGTGGAGCTGACGAAGAGCCTGCT 1113
 QY 1120 GGTGCCACCATTTGGGAACCGATGACATCAAAAGGAAACAAAGCCGGTGGCCAGC 1179
 Db 1114 GCGCTTGGCAATACACCGCAAGGACGCGCGGTCGGCAGCACCATCCCGACCCGCTC 1173
 QY 1180 GACCCCTCTATTCGCCACACCCGATCATGACCGATGCGGATATGGCGATAAGGTAAT 1239
 Db 1174 GCGGCGCC---AGGCGCGTCCCGACGATGTGGCCACTGACCTCTCGCTGCGGTGGAT 1230
 QY 1240 CCGACTATCGCGCTATCTGCGAAATTCATGCGCGGATCTGAGTACTTCAAGAAATCT 1299
 Db 1231 CCGATCTATGAGCGGATCAGCGTCTGCTGGTGGAAACACCCGAGGAATTCGCCACAG 1290
 QY 1300 TTCGGAAGGCGTGTCAAGCTGACGACCGTGGCTGACCTGGGCGCCGAATACAGTTACATC 1359
 Db 1291 TTCGGAAGGCGTGTCAAGCTGATCCACCGAGACATGGGTCCCGTTTCGAGATACCTT 1350
 QY 1360 GCGCGGAAGTCCCGGAGAGACCTGATTTGGCAAGACCCGATTCGCGCAGGTAACACC 1419
 Db 1351 GGGCGCGTGGTCCCGAAGCAGACCTGCTGTGGCAGGATCCGCTCCCTGCGCTACGCCAC 1410
 QY 1420 GACTACTCGAAGAAGTGG-----TCAAGCAGAAATTCGCAAAAGTGGCCTG 1467
 Db 1411 GACCTCTGGCGAAGCCGAGATTCGACGCTTAAGAGCCAGATCCGGCATTCGGGATG 1470
 QY 1468 GACATATGATGATGCTCTCCACCGCTTGGGACAGTGGCCGCTACTATTCGCGGTTCCGAT 1527
 Db 1471 ACTGTCTCACAGCTAGTTTCGACCGCATGGGCGGCGCTGCTGCTCGTGTGATAGCAG 1530
 QY 1528 ATGCGCGGCGGTGCTAACGGTGGCGCGATTCGCTTGGCCCCACAGAACAGTGGCAGGCG 1587

Db 1531 AAGCGCGCGCGCAACCGGTGGTCCGATCCGCTGCGAGCCACAAAGTCCGCTGGAGGTC 1590
 QY 1588 AACGAGCCGGA---GCGCTTGGCGAAAGTGTGAGCGCTCTAGGAGCAGATCT----- 1636
 Db 1591 AACGAGCCCGAGCGGGATCTCGGCAAGGTCTATTCGACCCCTGGAAGAGATCCAGAGTCA 1650
 QY 1637 -----CTGCCGACACCGCGGCTAGCATCGGGAGGATGATGCTGCTGCGC 1680
 Db 1651 TTCAACTCCGCGGCGCGGGAACATCAAGTGTCTTCGCGGACCTCGTCTGCTCGT 1710
 QY 1681 GGTAGCTTAGCATCGAAGAGCCGGAAGCAGCAGGTACGATGTCGGGCTTCCCTTC 1740
 Db 1711 GGCTGTGCGCATAGAAAGCAGCAAGGCGGCTGGCCACAACATCACGCTGCCCTTC 1770
 QY 1741 CTGAAGGCGCTGGCGATGCGACCGCGGATGACCGAGCGAGATCTCTTCACACCGCTG 1800
 Db 1771 ACCCGGCGCGCACGATGCGTCGCGAGCAACACCGAGCTGATCTTTGCCGCTG 1830
 QY 1801 GAGCGCTGGCGGATGGCTTCGCAACTGGCGAAGAGAGATGTTGGTGAAGCCGGA 1860
 Db 1831 GAGCCAAAGCGAGATGGCTTCGAAACTACCTCGGAAAGGCAACCCGTTGCCGCGAG 1890
 QY 1861 GAGATGCTGCTGGATCGTGGCAGCTGATGGGCTTAACCGGCGCCGAAATGACCGTCTG 1920
 Db 1891 TACATGCTGCTCGAAGGCGAAACCTGCTTACGCTCAGTGGCCCTGAGATGACGCTG 1950
 QY 1921 CTGGCGCTATGCGGCTACTGGGCAACCACTATGTTGGCACCACCAACAGCGCTATTAC 1980
 Db 1951 GTAGTGGCTGCGGCTCTCGCGCAAACTACAAGCGCTTACCGCTGGCGTGTTCACC 2010
 QY 1981 GATTGTGAAGCGCATGACCAACGACTTTTGTGTAACCTGACGATATGGGGAACAGC 2040
 Db 2011 GAGGCTCCGAGTCTACTGACCAACGACTTCTTCGTGAACCTGCTCGACATGGGTATCAC 2070
 QY 2041 TGGAGC---CGTAGGTAGCAACCGCTACGNAATCCGCGACCGCAAGACCGTGGCGTG 2097
 Db 2071 TGGAGCGCTTCCGAGAGATGACGGGACCTACAGGGCAAGGATGGCAGTGGCAAGTG 2130
 QY 2098 AAGTGACCGCTCCGCGGTGGATGTTGTTTGTTCCTCAACTCGCTACTGCGCTCTTAC 2157
 Db 2131 AAGTGACCGCGAGCGCGTGGACCTGTTCTTCGGGTCCAACCTCGAGCTTGGCGGCTT 2190
 QY 2158 CGAAGTGTAGCGCCAGCAGATTAACGCGAGAGTTCGTGTCAGAGCTTCGTCGCGCC 2217
 Db 2191 CTCGAGGCTATGCGCGGATGACGCGAGCGAAGTTCGTGTCAGGACTTCGTCGCTGCC 2250
 QY 2218 TGGACCAAGTATCAACCGCGACCGTTTCGACGT 2252
 Db 2251 TGGGCAAGGTATGAACCTCGACAGGTTCGACGT 2285

RESULT 8

AAAT90400
 ID AAAT90400 standard; DNA; 2331 BP.
 XX
 AC AAAT90400;
 XX
 DT 25-MAR-2003 (updated)
 DT 21-JAN-1998 (first entry)
 XX
 DE Mycobacterium tuberculosis wild-type katG gene.
 XX
 KW Isoniazid resistance; restriction fragment length polymorphism;
 KW RFLP; katG gene; catalase; peroxidase; multiple drug resistance;
 KW Isonicotinic acid hydrazide; INH; tuberculosis; antibiotic;
 KW tuberculostatic; ss.
 XX
 OS Mycobacterium tuberculosis.
 XX
 FH Key
 FT CDS
 FT 70..2292
 FT /*tag= a

Qy	1360	GGCCCGGAGATGCGCGGAGAAAGACCTGATTTGGCAAGACCCGATTCGCGCAGGTAAACAC	14119
Db	1351	GGGCGCGTGTGTCGCCCAAGCAGACCCCTGCTGTGCAGGATCCGGTCCCTGCGGTACGCCAC	14110
Qy	1420	GACTACTGCCAANAAGTGG-----TCAACGACAGAAATTCGACAAAGTGGCCCTG	1467
Db	1411	GACTCTGTCGGGAGAGCCGAGATTGCCAGCCTTAAGAGCCAGATCCGGGCATCCGGGATTG	1470
Qy	1468	AGCATTTAGTCAGATGTCTCCACCGCTTGGGACAGTCCCGCTACTTTATCGCGTTCCCGAT	1527
Db	1471	ACTGTCTCACAGCTAGTTTTCGACCGCATGGCGCGCGTCTGCTTCCGTGTGAGGCAC	1530
Qy	1528	ATGGCGGGCGGTGCTAACGTTGCCCGGATTCGCTTGGCCCCCACAAGACGAGTGGCAGGGC	1587
Db	1531	AAGCGGGCGCGCCCAACGTTGTCTCGCTCCGCTCGACCCAAAGTCGGGTGGGAGTCT	1590
Qy	1588	AACGAGCCGGA--GCGCCTGGGGAAGTCTGAGCGTCTACGACGAGATCT-----	1636
Db	1591	AACGACCCACGGGAGATCTCGCCAGAGTCAATTCGCACCTCGGAAGAGATCCAGGAGTCA	1650
Qy	1637	-----CTGCCGACACCGCGCTAGCATCCGCGACGCTGATCGCTTCCTGGCC	1680
Db	1651	TTCAACTCCGCGCGCGGGGAACATCAAGTGTCTTCGCCGACCTCTGCTGCTCGGT	1710
Qy	1681	GGTAGCGTAGGCATTCGAGAAAGCGGGAAGCAGCAGGTTACGATGTGCGCGTTCCTTC	1740
Db	1711	GGCTGTGCCGCCATAGAGAAAGCAGCAAGAGCGGCTGGCCACAACATCACGGTGCCCTTC	1770
Qy	1741	CTGAAAGGCGGTGGCGATGCGACCGCGCGAGATGACCGACGACAGTCTTCGCGACCGCTG	1800
Db	1771	ACCCGCGCCGACGGATGCGTTCGAGGAACAACCGACGTGGAATCTCTTGGCGCTGCTG	1830
Qy	1801	GAGCGGTGGCCGATGGCTTCCGCAACTGGCAGAAAGAGATGTGTTGTAAGACGGAA	1860
Db	1831	GAGCCCAAGCAGATGGCTTCGGAACACTACCTCGGAAGGCAACCGCTTGC CGCGCGAG	1890
Qy	1861	GAGATGTGCTGGATCTGCGCAGCTGATGGGCTTAACCGGCCGGAATGACCGTGTCTG	1920
Db	1891	TACATGTGCTGCACAAAGCGCAACCTGCTTACGCTCAGTCCGCTGAGATGACGGTGTCTG	1950
Qy	1921	CTGGGCGGTATGCGCTACTGGGCACCAACTATGTTGGGCACCAACACGCGCTATTCAC	1980
Db	1951	GTAGTGGCCTGGCGTCTTCGGGCAAACTACAAGCGCTTACCGTGGCGGTGTTCACC	2010
Qy	1981	GATTTGTAAGGCCAGTTGTACCAACGACTTTTTTTGTGAACCTGACCGATATGGGGAAC	2040
Db	2011	GAGCCCTCCGAGTCACTGACCAACGACTTCTTCGTGAACCTGCTCGACATGGGTATCAC	2070
Qy	2041	TGGNAGC--CGGTAGTAGCAAGCGCTACGAATTCGGGACCGCAAGACCGGTGCGGTG	2097
Db	2071	TGGAGCGCTCGCCAGCAGATGACGGGACCTACACGGGCAAGGATGCGATGGCAAGTGT	2130
Qy	2098	AAGTGAACCGCCTCGCGGTTGGATCTGGTATTTGGTTTCAACTCGCTACTGCGCTCTTAC	2157
Db	2131	AAGTGAACCGCGACCGCGTGGACCTGGTCTTCGGGTCCAACTCGGAGTTGCGGGCGCTT	2190
Qy	2158	GCAGAAGTGTACGCCACAGGACGATAACGGCGGAGAAGTTTCGTCAAGACTTCGTGCGGCC	2217
Db	2191	GTGAGGTCTATGGCGCCGATGACGCGCACCGCAAGTTCGTGCAGGACTTCGTGCTGCTGCC	2250
Qy	2218	TGGACCAAGTGATGAACGCCGACCGTTTCGAGCT	2252
Db	2251	TGGGACAAGGTGATGAACCTCGACAGCTTCGAGCT	2285

RESULT 9
AAX00817
ID AAX00817 standard; DNA; 2331 BP.
XX
XX AAX00817;
XX
DT 26-MAR-1999 (first entry)

	M. tuberculosis catalase peroxide (katG) gene sequence.
XX	Catalase-peroxidase; katG; Isonicotinic acid hydrazide; tuberculosis;
DE	INH; diagnosis; detection; ss.
KW	
KX	Mycobacterium tuberculosis.
OS	
XX	
Key	Location/Qualifiers
FH	70..2292
CDS	/tag= a
FT	/gene= "katG"
FT	/product= "catalase peroxide"
FT	/note= "the start codon is not indicated"
XX	
PN	WO9850585-A1.
XX	
PD	12-NOV-1998.
XX	
PF	06-MAY-1998; 98WO-US09285.
XX	
PR	07-MAY-1997; 97US-0852219.
XX	
PA	(MAYO-) MAYO FOUNDATION.
XX	
PI	Cockerill FR, Kline BC, Uhl JR;
XX	
DR	WPI: 1999-070099/06.
DR	P-PSDB: AAW95398.
XX	
PT	Detection of Mycobacterium tuberculosis - by amplifying katG gene
PT	and detecting specific fragment, and optionally identifying
PT	INH-resistant strains by detecting specific mutation
XX	
PS	Claim 3; Fig 7; 83pp; English.
CC	The invention relates to a novel method of detecting Mycobacterium
XX	tuberculosis. The method comprises amplifying the DNA in the samples to
CC	generate a detectable amount of amplified DNA comprising a catalase-
CC	peroxide (katG) DNA fragment with sequence of bases 904-1523 of the
CC	M. tuberculosis strain H37Rv MC katG gene sequence. The method optionally
CC	further comprises determining if the katG DNA fragment has a serine to
CC	threonine mutation in codon 315 (S315P mutation), indicative of an
CC	Isonicotinic acid hydrazide (INH)-resistant M. tuberculosis strain. The
CC	method can be used to detect M. tuberculosis in biological fluids,
CC	specially human sputum, useful to diagnose tuberculosis. This disease is
CC	a major cause of human morbidity and mortality, and conclusive diagnosis
CC	and subsequent treatment depends on identification of the etiologic agent
CC	M. tuberculosis. INH has been used in tuberculosis treatment, but
CC	INH-resistant strains have emerged; the method allows such drug-resistant
CC	strains to be identified. The present sequence represents the wild-type
CC	M. tuberculosis katG gene sequence.
XX	
SQ	Sequence 2331 BP; 455 A; 734 C; 763 G; 379 T; 0 other;
	Query Match 27.3%; Score 618.6; DB 20; Length 2331;
	Best Local Similarity 58.7%; Pred. No. 3.4e-165;
	Matches 1253; Conservative 0; Mismatches 819; Indels 63; Gaps 8
Oy	160 AACAAAGATTGGTGC CGGAAGGT TTGAACCTGGATATTTTGCA TCAGAAGATCGCAA 219
Dd	172 AACCAGGACTGGTGC CCACA CGCGGTCA ATCTGA AGGTACTGCCA CAACAAACCGCGC GTC 231
OY	220 TCAGACCGGANGGATTC GGATTTCA ACTACCGT GAAGAAGTA GCAGACTCGATTTCG AC 279
Dd	232 GCTGACCCGATGGGTGG CGGTTTG CATAT GCCGCGGAGGT CGCACCATCGACGT TGAC 291
OY	280 GCGCTGAAGAAGATGT CCACGCGT TGATG ACCGATG CCGAACAGAGTGGTGG CCCCTGCAC 339
Dd	292 GCCCTGACGCGGACATCG AAGTAGTATG ACCACCTCGC AGCCGTGGTGGCCCCGCCGC AC 351
OY	340 TGGGGGCACTACGCGGTT TGATGATCCGTATGG CTTGGGCATTCGGCTGGCACCTACCGT 399

Db 352 TACGGCCACTACGGCCGCTGTTTATCCGGATGGCGTGGACGCTGCCGGACCTACCGC 411
Qy 400 ATTGCTGATGGCCGTGGGGCGGTGGTACCGGAAGCAGCGCTTTGACCGCTCAACTCC 459
Db 412 ATCCACAGCGCGCGCGCGCGCGCGGCGGATGACAGCGTTTCGGCGCGCTTAACAGC 471
Qy 460 TGGCCGACAACTGACCGCTGGATTAAGAGCGCGCGCTGCTGTGGCGCGATCAAGAAG 519
Db 472 TGGCCGACAACTGACCGCTGGATTAAGAGCGCGCGCGCTGCTGTGGCGCGTCAAGAAG 531
Qy 520 TACGGCAACAAATACGCTGGCAGACCTGATGATTTCTGGCTGGCAGCGGTGGCTATGAG 579
Db 532 TACGGCAAGAACTGCTATGGCGGACCTGATGTTTTCGGCGCACTCGCGCTGGAA 591
Qy 580 TCCATGGGCTTACCTGCTTACGGCTTCTTTTCGGCGCGCTGATATTTGGGAACCGGAA 639
Db 592 TCGATGGCTTCAAGAGCTTCGGCTTCGGCTTCGGCGGCTCGACCACTGGGAGCGCG-- 649
Qy 640 AAGATATCTACTGGGTGACGAAAGAGTGGTGGCAGCTTCTGACGAACGCTACGGC 699
Db 650 -ATGAGGTCTATTGGGSCAAGAACCCACTGGCTCG- - - - -GCGATGAGCGTTACAGC 702
Qy 700 GAGCTGAACAGCAGACCATGGAAACCCGCTGGCGCTGTCCAAATGGTCTGATC 759
Db 703 GGTAAAGCGGATC- - - - -TGGAGAACCCGCTGGCCCGGTGCGAGATGGGCTGATC 753
Qy 760 TATGTGAACCCGGAAGGTGTTAACGGCCACCCTGATCCGCTGAGAACCGCACAGCGTA 819
Db 754 TAGCTGAACCCGGAAGGTGTTAACGGCCACCCTGATCCGCTGAGAACCGCACAGCGTA 813
Qy 820 CTTGAAACCTTCGCCGTATGGCGATGAACGAGCAAAACCGAGCCCTCACAGCTGGC 879
Db 814 CGCGAGACGTTTCGGCGCATGGCCATGAACGAGCTCGAAGACAGCGCGCTGATCGTCGC 873
Qy 880 GGCACACCTCGGTAAATTCACGCTAATGGCAATGCCCTCGCTTAGCCCTGACCCCA 939
Db 874 GGTCAACCTTCGGTAAGACCATGGCGCGCGCGCGCGCATGTGGTGGCCCGGAACCC 933
Qy 940 AAGGCTCTGACGTTGAAACACGAGGCTTAGTTGGGGCAACCCCAACATGACAGGCAAG 999
Db 934 GAGGCTGCTCCGTGAGCAGATGGGCTTGGGCTGGAAGAGCTCGTATGCACCGGAAAC 993
Qy 1000 GCAAGCAACCCCTGACCTGGGTATCGAAGGTGCTTGGACCAACCAACCCCAAGAAATTC 1059
Db 994 GGTAAAGACCGCATCACAGCGCATCGAGTGTATGGACGAACCCCGCGAAGATGG 1053
Qy 1060 GATATGGGCTATTTCGACCTGCTGTTCCGGCTACAAATTTGGGAAGTGAAGAGCTCTGCC 1119
Db 1054 GACAACAGTTTCTCGAGATCTGTACGGCTACGAGTGGAGCTGACGAAGACCTGCT 1113
Qy 1120 GGTGCCACCATTTGGGACCGATGACATCAAAAGAAACAAAGCGGTTGACGCCAGC 1179
Db 1114 GCGGCTTGGCAATACACCGCAAGGACGGCGCGGTGCCGGCACTATCCCGGACCGGTT 1173
Qy 1180 GACCCCTCTATTCCGACACCCGATCATCACCGATGCGGATGCGGATTAAGCTAAAT 1239
Db 1174 GCGGCGC- - - - -AGGGCGCTCCCGACGATGCTGGCCACTGACCTCTCGCTGCGGGGTGAT 1230
Qy 1240 CCGACCTATCGGCTATTCGCAAAATTCATGGCGCATCTGAGTACTTCAAGAAACT 1299
Db 1231 CCGATCTATGAGCGGATACCGCTCGCTGGCTGGACACCCCGAGGANTTGGCCGAGAG 1290
Qy 1300 TTCGGAAGCGGTGTTCAAGCTGACGACCGCTGACCTGGGCGCGAAATCACCTTATATC 1359
Db 1291 TTCGCAAGCGCTGTACAGCTGTATCCACCGAGACATGGTCCCGTTGGAGATACCTT 1350
Qy 1360 GGGCGGAAGTGGCGGACAGACCTGATTTGGCAAGACCCGATTCGGCAGGTATACACC 1419
Db 1351 GGGCGCTGTGTCACCAAGCAGACCCCTGCTGTGACAGATCCCGTCCCTGCGGTACGCCAC 1410
Qy 1420 GACTACTGCGAAGAGTGG- - - - -TCAAGCAGAAATTCACAAAGTGGCCTG 1467
Db 1411 GACCTGTCGGCGAAGCCGAGATTCGACCGCTTAAGACCGAGATCCGGGATCGGATG 1470

Qy 1468 AGCATTAAGTGAAGTGTCTCCACCGTGGACAGATGCCCGTACTTATTCGCGGTTCGCAT 1527
Db 1471 ACTGTCTCACAGCTAGTTTCGACCGCATGGCGGCGTGGTTCGCTCGGTAGGCAC 1530
Qy 1528 ATGCGCGCGCGTGTAAACGGTGGCCCGATTCGGTTFGGCCCCACAGAACAGTGGCAGGGC 1587
Db 1531 AAGCGCGCGCGCGCAACCGTGGTTCGATCCGCTGACAGCAAGTTCGGTGGGAGTTC 1590
Qy 1588 AAGGAGCGGGA- - -GCGCCTGGGGAAGTGTGAGCGTCTACGAGCAGATCT- - - - - 1636
Db 1591 AAGACCCCGACGGGATCTGGCAAGGTCAATTCGCACCTTGAAGAGATCCAGGAGTCA 1650
Qy 1637 - - - - -CTGCGCACCGCGGCTAGCATCGCGACGTGATCTTCTGTCGCC 1680
Db 1651 TTNACCTCCGCGCGCGGGAACATCAAGTGTCTTCGCGCACTCTGTCGCTCGGT 1710
Qy 1681 GGTAGCTAGGATCGAGAAAGCGGCAAGAGCAGCAGGTTACGATGTGCGGTTCCCTTC 1740
Db 1711 GGCTGTGCCCGCATAGAGAAAGCAGCAAGCGCGCTGGCCACACATCACGGTGGCCTTC 1770
Qy 1741 CTGAAAGCGCGTGGGATGCGCGCGAGATGACCGACGACGACTCTCTTCGACCGCTG 1800
Db 1771 ACCCGCGCGCAGGATGCGTCGAGGAACAAACCGACGTGGAATCTTTTGGCGTCTG 1830
Qy 1801 GAGCGGCTGCGCATGGCTTCGCAACTGGCAGAAAGAGATGTGTGTAAGCCGGAA 1860
Db 1831 GAGCCCAAGCAGATGGCTTCGGAACACTACCTCGGAAGGGCAACCGGTTGCGCGCGAG 1890
Qy 1861 GAGATGCTGTGATCGTGGCAGCTGATGGGCTTAAACCGCGCGGAAATGACCGTGTG 1920
Db 1891 TACATGCTGTGACAAAGCGAACCTGTCTTACGCTCAGTCCGCTGAGATGACGGTGTG 1950
Qy 1921 CTGGGCGTATGCGCTACTGGCACAACACTATGTTGGCACAACACACGCGCTATTAC 1980
Db 1951 GTAGTGGCTGCGCTCTCGCGCAACACTCAAGCGCTTACCGCTGGCGGTGTTTAC 2010
Qy 1981 GATTTGAAGCGCAGTTGACCAACGACTTTTTTGTGAACCTGACCGATATGGGGAACAGC 2040
Db 2011 GAGGCTCCGAGTCACTGACCAACGACTTCTCTGTGAACCTGCTCGACATGGGTATCAC 2070
Qy 2041 TGGAGC- - -CGTAGGTAGCAACCGCTAGCAAAATCGCGACCGCAAGACCGGTGCGGTG 2097
Db 2071 TGGAGCGCTCGCCACGATGACGGGACCTTACAGGGCAAGGATGGCAGTGGCAAGGTG 2130
Qy 2098 AAGTGACCGCTCGCGGTGGATCTGTTTGGTTCCAACTGCTACTGCGCTCTTAC 2157
Db 2131 AAGTGGACCGCAGCGCTGGACCTGGTCTTCGGGTCCAACTCGGAGTTGCGGGCGCTT 2190
Qy 2158 GCAGAAAGTGTACGCCAGGACGATAACGGCGAAGTTCGTCAAGACTTCTGTCGCGGCC 2217
Db 2191 GTCGAGTCTATGGCGCGATGACGCGACGCCGGAAGTTCGTGCGAGACTTCTGTCGCTGCC 2250
Qy 2218 TGGACCAAGTGTATGAACCGCGACCGTTTCGAGCT 2252
Db 2251 TGGGACAAGTGTGAACCTCGACAGGTTTCGAGCT 2285

RESULT 10

ABX09140/c
ID ABX09140 standard; DNA; 82993 BP.

XX
AC ABX09140;

XX
DT 08-APR-2003 (first entry)

XX
DE Mycobacterium tuberculosis H37Rv BAC clone BAC-Rv221.

XX
KW Mycobacteriosis; survival; virulence; protective antigen; vaccine;

XX
KW Mycobacterial disease; tuberculosis; leprosy; ds; cosmid.

XX
OS Mycobacterium tuberculosis.

XX

38857 GGCTGTGCCGCATAGAGAAAGCAGCAAGAGCGGCTGCCACACATCAGCGTCCCTTC 38798

1741 CTGAAGGCGGTGGCGATGCGACCGCGGAGATGACCGACGAGACTCCTTCGCCACCGGTG 1800

38797 ACCCGGCGCCACGAGTGCCTGCGAGGAACAAACCGACGTTGAATCCTTTGCCGTGCTG 38738

1801 GAGCGCTGGCGATGCTGCGCACTGCGAGGAAGAGATGATGGTGAAGCCGGAA 1860

38737 GAGCCGAAGCGAGATGCTTCCGAACACTACCTCGGAAGGCAACCCGTTGCCGCGGAG 38678

1861 GAGATGCTGCTGGATGCTGCGCAGCTGATGGCTTAACCGCGCGGAAATACCGTGTG 1920

38677 TACATGCTGCTCGACAGGCGAAGCTGCTTACGCTCAGTGCCTCGATGACCGTGTG 38618

1921 CTGGGCGGTATGCGCGTACTGCGGCAACCACTATGTTGGCAACCAACGCGGTATTCACC 1980

38617 GTAGTGGCTTGGCGCTCCTCGGCGCAACCTACAAAGCGCTTACCGCTGGCGGTGTTACC 38558

1981 GATTGTGAAGGCCAGTTGACCAAGCACTTTTGTGAACCTGACCGGATATGGGAACAGC 2040

38557 GAGCCCTCCGAGTCACTGACCAACGACTTCTTCTGTAACCTGCTCGACATGGGTATCACC 38498

2041 TGAAGC---CGTAGTAGCAACCGCTACGAATCCGCGACCGCAAGACCGTGGCGTG 2097

38497 TGGAGCCCTCGCAGCAGATGACGGGACCTACCGGCGAAGATGCGAGTGGCAAGTG 38438

2098 AAGTGGACCGCTCGCGGTGGATGCTGGTATTTGGTTCCAACTCGCTACTCGCTCTTAC 2157

38437 AAGTGGACCGCGACCGTGGAGCTGCTTCCGGTCCAACTCGGAGTTGCGGCGCTT 38378

2158 GCAGAGTGTACGCCCGGAGGATACCGGCGGAGAGTTGCTCAGAGACTTGTGCGCGGCC 2217

38377 GTCAGGTCTATGGCGCGGATGACGCGCAGCGCAAGTTGCTGCGAGGACTTGTGCGCTGCC 38318

2218 TGGACCAAGTGATGAACCGCGCGCTTTCGACGT 2252

38317 TGGACCAAGTGATGAACCTCGACAGTTGACGT 38283

RESULT 11

AA199683/C

ID AA199683 standard; DNA; 4403765 BP.

XX

AC AA199683;

XX

DT 15-JAN-2002 (first entry)

XX

DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.

XX

XX Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;

KW variation; epidemiology; patient treatment; epidemic monitoring; ds.

XX

XX Mycobacterium tuberculosis.

OS

XX

XX US6294328-B1.

PN

XX

PD 25-SEP-2001.

XX

XX 24-JUN-1998; 98US-0103840.

XX

XX 24-JUN-1998; 98US-0103840.

XX

XX (GENO-) INST GENOMIC RES.

PA

XX

XX Fleischmann RD, White OR, Fraser CM, Venter JC;

XX

XX WPI; 2001-647261/74.

DR

XX

XX Evaluating strain variation of Mycobacterium tuberculosis, comprises

PT determining the nucleotide sequence of the strain at positions in the

PT genome corresponding to positions where M. tuberculosis strains CDC

PT 1551 and H37Rv differ

XX

Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.

The invention relates to evaluating strain variation within and between different populations of the tuberculosis or related Mycobacterium by determining the Mycobacterium tuberculosis or related Mycobacterium by determining the nucleotide sequence of the first strain at positions that differ in the complete sequence of the genome that correspond to positions that differ in the complete nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and H37Rv (AA199682). The method is useful for evaluating strain variation of M. tuberculosis and has valuable application in the fields of M. tuberculosis genetics, epidemiology, patient treatment and epidemic monitoring.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=6294328B1.

Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

Query Match 27.3%; Score 618.6; DB 22; Length 4403765;
Best Local Similarity 58.7%; Pred. No. 1.2e-163;
Matches 1253; Conservative 0; Mismatches 819; Indels 63; Gaps 8;

QY 160 AACAAAGATTGGTGGCGGGAAGGTTGAACCTGGATATTTTGCATCAGCAAGATCGCAAA 219

DB 2153267 AACAGAGACTGGTGGCGCAACCGGCTCAATCTGAAGGTACTGCACCAAAACCCGCGCTC 2153208

QY 220 TCAGACCCGATGGATCGGGATTTCAACTACCGTGAAGAAGTACGAAGCTCGATTTCGAC 279

DB 2153207 GCTACCCGATGGTGGCGCTTCGACTATCCCGGAGGTTCGGACCATCAGCTTGAC 2153148

QY 280 GCGTGAAGAAAGTATCCAGCGCTTGATGACCATAGCCAAAGATGGTGGCGCGCTGAC 339

DB 2153147 GCCCTGACGGGACATCGAGGAAGTATGATACCACCTCGCAGCGGTGGTGGCGCGCGAC 2153088

QY 340 TGGGGGCACTACCGCGGTTTCATGATCGTATGGCTTGGCACTCCGCTGGCACTACCGT 399

DB 2153087 TACGGCCACTACGGCGCGCTGTTATCCGATGGCTGGCAGCTGGCGCGCACTACCGC 2153028

QY 400 ATTGCTGATGGCGCTGGGGCGGTGTTACCGGAAGCCAGCGCTTTCGACCCCTCAACTCC 459

DB 2153027 ATCCACGACG 2152968

QY 460 TGGCGGCAACAGTCAAGCTGGATTAAGCGCGCGCTGTGTGTGGCGCGATCAAGAAGAG 519

DB 2152967 TGGCGGCAACAGTCAAGCTGGATTAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2152908

QY 520 TACGCAACAAATCAGTGGCAGACTGATGATTTGGTGGCAGCGCTGGCTGGCTTATGAG 579

DB 2152907 TACGCAACAAATCAGTGGCAGACTGATGATTTGGTGGCAGCGCTGGCTGGCTTATGAG 2152848

QY 580 TCCATGGGCTTACCTGCTTACGGCTTCTCTTTTCGCGCGCGCTCGATATTTGGAAACCGCAA 639

DB 2152847 TCGATGGCTTCAAGACGTTTCGGGTTTCGGCTTCGCGCGCGCTCGACAGTGGAGCCGC 2152790

QY 640 AAGATATCTACTGGGTGACGAAAGAGTGGTGGCACTTCTGACGAACTGCTACGCG 699

DB 2152789 -ATGAGTCTATTGGGCAAGAAAGCACTGGCTGCG-----CGATGAGCGTTACAGC 2152737

QY 700 GACGTGAACAGCCAGACCATGGAACACCGCTGGCGCGCTGCGTGGCTGCG-----CGATGAGCGTTACAGC 2152737

DB 2152736 GGTAAAGCGGATC-----TGGAGAACCGCTGGCGCGCTGCGATGGGCTGATC 2152686

QY 760 TATGTGAACCGGAAAGTGTAAACGGCACTTACCGGCACTTATCGGCTGAGAACCGCACAGGTA 819

DB 2152685 TACGTGAACCGGAGGCGCGCAAGCGCAACCGGCACTTACCGGCACTTATCGGCTGAGAACCGCACAGGTA 819

QY 820 CTTGAAACCTTGGCGCGCTATGGCGATGAACGACGAAAGAAACCGGCGCGCTTACAGCTGGC 879

DB 2152625 CCGGAGACGTTTGGCGCGATGGCGATGAACGAGCTGCAAGAGCGGCGCTGATCGTGGC 2152566

QY 880 GGCACACCGCTCGGTAAATGTCAGGTAATGGCAATGCTTGGTGGCTAGCCCTGACCA 939

DB 2152565 GGTACACACTTTCGTAAGACCCATGGCGCGCGCGCGCGCTGCTGGTGGCGCGCGCGCGCGCG 2152506

```
Qy 940 AAGCCCTCTGACGTTGAAACACAGGCGTTAGTTGGGGCAACCCCAACATGCAAGGGAAG 999
Db 2152505 GAGGCTCTCGCTGGAGCAGATGGGCTGGGCTGGAAGAGCTGTTATGGCACCGGAACC 2152446
Qy 1000 GCAAGCAACCGGTGACCTCGGCTGATGAAAGGTCTTTGGACCAACCAACCCCAACAAATTC 1059
Db 2152445 GGTAAAGACGGGATCACACAGGGGATCGAGGTCTATGGAGCAACCCCGCAAGAAATGG 2152386
Qy 1060 GATATGGGCTATTTTCGACCTGCTCTTCGGCTTACAAATTTGGAACTGAAAAAGAGTCCTGCC 1119
Db 2152385 GACAACAGTTTCTCTCGAGATCTCTACGCTACGAGTGGGAGCTGACGAAGAGCCCTGCT 2152326
Qy 1120 GGTGCCCAACATTTGGGAACCGATTGACATCAAAAAGGAANAACAGCCGGTTGAGCCGACG 1179
Db 2152325 GGCCTTTGGCAATACACGCCAAGAGCGGCGCGGTGCGCGACCATCCCGGACCCGCTTC 2152266
Qy 1180 GACCCCTCTATTCCGCCACAACCCCATCATGACCCGATCGGATATGGCGATAAAGGTAAT 1239
Db 2152265 GCGCGGCC---AGGCGCGTCCCGCAGCATGCTGGCCACTGACCTCTCGCTCGGCTGGAT 2152209
Qy 1240 CCGACCTATCGCGTATCTCGGAAAAAATTCATGGCCGATCTGAGTACTTCAAGAAAAACT 1299
Db 2152208 CCGATCTATGAGCGGATCACGCGTCTGCTGGCAACCCCGAGGAATTGGCCGACGAG 2152149
Qy 1300 TTCGGAAGGGGTGTTCAAGCTGACGCAACCGTGACCTGGCCCGGCAAAATCAGTTACATC 1359
Db 2152148 TTCGCCAAGGCTGGTCAAGCTGATCCACCGAGACATGGTCCGCTTCCGAGATACCTT 2152089
Qy 1360 GGGCCGGAAGTCCGGCAGAGAGACTGATTTGGCAAGACCCGATTCGGCAGGTAAACAC 1419
Db 2152088 GGGCGGTGTCGCCCAAGCAGACCTGCTGTGCGAGGATCCGGTCCCGGTGAGCCAC 2152029
Qy 1420 GACTACTCCGAAGAAGTGG-----TCAAGCAGAAAAATTGCACAAAGTGGCGTG 1467
Db 2152028 GACCTCTGCGGAAGCGGAGATTGCCAGCTTAAAGCCAGATCCGGGCATCCGGATTG 2151969
Qy 1468 AGCATATGATGAGTGTCTCACCGCTTGGGACAGTCCCGCTACTTATFCGGGTTCGGAT 1527
Db 2151968 ACTGCTCTCACAGCTAGTTTTCACCGCATGGCGCGCGCTGCTGCTTCCGTGGTAGGCAC 2151909
Qy 1528 ATGCGCGCGTGTAAAGGTGCGCGCATTCGCTTGGCCCAACAGACGAGTGGCAGGC 1587
Db 2151908 AAGCGCGCGCGCCALAGGTGTGTCATCCGCTCGACGCCACAGATCGGGTGGGAGTCT 2151849
Qy 1588 AACGAGCCGGA---GCGCCTGGCGAAAGTGTGAGCGTCTACGAGCAGATCT----- 1636
Db 2151848 AACGACCCGACGGGGATCTGCGCAAGTCAATTCGCAACCTGGAGAGATCCAGAGTCA 2151789
Qy 1637 -----CTGCCGACACCGCGCTAGCTATCGGGGAGCTGATCGTTGCGCC 1680
Db 2151788 TTCAACTCCGCGGCGCGGGGAACATCAAAAGTGTCTTCCGCGACCTCGTCTCGTCCGT 2151729
Qy 1681 GGTAGCGTAGCATCGAAGACCGCGAAGACGACAGTTCAGATGTGCGGTTCCTTC 1740
Db 2151728 GGCTGTGCCGCCATAGAGAAGAGAGAAAGCGGCTGCGCCACAACATCACGGTGCCTTC 2151669
Qy 1741 CTGAAAGCCGTGGCGATGCGACCGCGAGATGACCGACGAGACTCTTCCGACCGCTG 1800
Db 2151668 ACCCGGGCGCACGGATGCTGCGCAGGAACAACCGACGTGGATCTTTGCGGTGCTG 2151609
Qy 1801 GAGCGGTGGCGATGGCTTCGCAACTGGCAGAAGAAGATGATGGTGAAGCCGGAA 1860
Db 2151608 GAGCCCAAGGCAGATGGCTTCGGAACACTACCTCGGAAAGGCAACCCGTTGCGGCCGAG 2151549
Qy 1861 GAGATGCTGTGGATGCTGCGACCTGATGGCTTAACCGCGCCGGAAATCACGCTGCTG 1920
Db 2151548 TACATGCTGTGCAAGGGCAACCTGCTTACGCTCAGTGCCTCGATGAGATACCGTCTG 2151489
Qy 1921 CTGGCGCGTATGCGCGTACTGGGCAACCACTATGGTGGCAACCAACAGCGGCTATTAC 1980
Db 2151488 GTAGTGGCTTGGCGCTCTCGGCGCAACTACAGCGCTTACCGCTGGGCGTGTTCACC 2151429
```

```
Qy 1981 GATTGTGAAGCCAGTTGACCAACGACTTTTTTGTGAACCTGACCGATATATGGGAACAGC 2040
Db 2151428 GAGCCCTCCAGTCACTGACCAACGACTTCTTCGTGAACCTGCTCGACATGGTATACACC 2151369
Qy 2041 TGGGAAGC---CGGTAGGTAGCAACGCTTACCAAAATCCGCGACCGCAAGACCGGTCCGCTG 2097
Db 2151368 TGGAGAGCCCTCGCCAGCAGATGACGGGACCTACCAAGGCAAGGATGGCAGTGGCAAGGTG 2151309
Qy 2098 AAGTGGACCGCTCGCGGGTGGATCTGTTATTTGTTTCCAACTCGCTACTCGCTCTTTAC 2157
Db 2151308 AAGTGGACCGCGACCGCGTGGACTGTTCTTCGGGTCCAACTCGGAGTTGGGGCGCTT 2151249
Qy 2158 GCAGAAGTGTACGCCAGGACGATAACGCGGAGAGTTTCGTACAGAGCTTCGTGCGCCGCC 2217
Db 2151248 GTCAGGTCTATGGCGCGGATGACGCGCAGCGAAGTTCTGTGCAGGACTTCGTGCTGCC 2151189
Qy 2218 TGGACCAAGGTGATGAACCTCGACGCTTTCGACGT 2252
Db 2151188 TGGACCAAGGTGATGAACCTCGACGCTTTCGACGT 2151154

RESULT 12
AAI99682/C
ID AAI99682 standard; DNA; 4411529 BP.
XX
AC AAI99682;
XX
DT 15-JAN-2002 (first entry)
XX
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN US6294328-B1.
XX
PD 25-SEP-2001.
XX
PF 24-JUN-1998; 98US-0103840.
XX
PR 24-JUN-1998; 98US-0103840.
XX
PA (GENO-) INST GENOMIC RES.
XX
PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX
PS WPI; 2001-647261/74.
XX
Evaluating strain variation of Mycobacterium tuberculosis, comprises
determining the nucleotide sequence of the strain at positions in the
genome corresponding to positions where M. tuberculosis strains CDC
1551 and H37Rv differ -
Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
XX
The invention relates to evaluating strain variation within and between
different populations of the tuberculosis bacterial pathogen,
Mycobacterium tuberculosis or related Mycobacterium by determining the
nucleotide sequence of the first strain at positions in the complete
sequence of the genome that correspond to positions that differ in the
nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
H37Rv (AAI99682). The method is useful for evaluating strain variation of
M. tuberculosis and has valuable application in the fields of
tuberculosis genetics, epidemiology, patient treatment and epidemic
monitoring.
Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?docID=6294328B1.
XX
Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;
SQ
```

Query Match		27.3%; Score 618.6; DB 22; Length 4411529;
Best local Similarity		58.7%; Pred. No. 1.2e-163;
Matches 1253; Conservative		0; Mismatches 819; Indels 63; Gaps 8;
QY	160	AACAAAGATGGTGGCGGAAGGTTGAACCTCGATATTTTGCATCAGCAAGATCGCAAA 219
DB	2156007	AACAGAGCTGGTGGCCCAACCGCTCAATCTGAAGTACTGCACCAAAACCCGGCGGTG 2155948
QY	220	TCAGCCGATGATCCGGATTTCAACTACCGTGAAGAAGTACGCAAGCTTCGATTTTCGAC 279
DB	2155947	GCTGACCGATGGTGGCGGTTCGACTATGCGCGGAGGTGCGCGACCATCGAGTTGAC 2155888
QY	280	GCCTGAAGAAAGATGTCACCGGTTGATGACCGATAGCCAAAGTGGTGGCCCGCTGAC 339
DB	2155887	GCCTGACCGGGACATCGAAGGAAGTGTGACCACTCGCGACCGCTGGTGGCCCGCGAC 2155828
QY	340	TGGGGGCACTACCGCGGTGTTGATGATCCGTTGGCATTGGCTTGGCACCCTCGCATACCGT 399
DB	2155827	TACGGCCACTACCGGCGCTGTTTATCCGGATGGCTGGCACCTGCGCGCACCTACCGC 2155768
QY	400	ATTGCTGATGGCGTGGGGCGGTGGTACCGGAAGCCAGCGCTTTGCACCGCTCAACTCC 459
DB	2155767	ATCCAGGACGGCGCGCGCGCGCGCGGCGCATCGACGGTTCGCGCGCTTAACAGC 2155708
QY	460	TGCGCGGACAACTGACCTGGATTAAGCGCGCGCTGCTGCTGGCGGATCAAGAAGAAG 519
DB	2155707	TGCGCGGACAACTGGATTTGGACAAAGCGCGCGCGCTGCTGCTGGCGGTTCAAGAAGAAG 2155648
QY	520	TAGCGCAACAAATACGCTGGGCGAGACCTGATGATTCTGGCTGGCAGCGGTGTTATGAG 579
DB	2155647	TAGCGCAAGAGCTCTATGGGCGGACCTGATTTGTTTTCGCGCGCACTGCGCGCTGGAA 2155588
QY	580	TCCATGGGTTACTCTTACGGCTTCTCTTTTCGCGCGCTGATATTTGGGAACCGGAA 639
DB	2155587	TCGATGGGTTCAAGACGTTTCGGCTTCGGCTTCGGCGCGGTCCAGCAGTGGGACCGC -- 2155530
QY	640	AAAGATATCTACTGGGGTGACGAAAGAGTGGCTGGCACCCTTCTCAGCAACCGCTAGGC 699
DB	2155529	-ATGAGGTCTATTGGGCAAGGAAGCAACCTGCTGCTG-----GCGATGAGCCTTACAGC 2155477
QY	700	GAGGTGAACAGCAGAGACCATGGAACCCGCTGGCGCTTCCAAATGGCTCTGATC 759
DB	2155476	GGTAACGGGATC-----TGGAGAACCCCTGCGCGGTGCAGATGGGGCTGATC 2155426
QY	760	TATGTGAACCGGAAGGTGTTAAGCGCACCTGATTCGCTGAGAACCGCACAGCAGTA 819
DB	2155425	TAGTGAACCGGAGGGCGGCAACCGGACCCGACCCGCGGCTGCGCGGCTGCGACATT 2155366
QY	820	CTTGAACCTTCGCCGTTATGGCGATGAACGCAAGCAACCCGCTCAGCTGCG 879
DB	2155365	CGGAGACGTTTCGGCGCATGGCCATGAACGAGCTGGAACAGCGGCTGATGCTCGGC 2155306
QY	880	GGCCACACGCTCGGTAAATGTCACGGTAATGGCAATGCTCTGCTTAGCCCTGACCCA 939
DB	2155305	GGTCACACTTCGGTAAGACCCATGGCGCGCGCGCGCGGATCTGGTCGCGCCCGGAACC 2155246
QY	940	AAAGCCTCTGAGCTTGAACACCGAGGCTTAGTGTGGGCAACCCCAACATGCAAGGCAAG 999
DB	2155245	GAGGCTGCTCCGCTGGAGCAGATGGCTTGGGCTGGAAGAGCTCGTATGCAACCGGAACC 2155186
QY	1000	GCNAGCAACCGCTGACCTCGGGTATCGAAGGTGCTTGACACCAACCAACCCAGCAATTC 1059
DB	2155185	GTTAAGGACGCGATACCAAGCGCATCGAGGTGATGGACGAACCCCGACGAATGG 2155126
QY	1060	GATATGGCTATTTCACCTGCTTGGCTACAAATTTGGAAGTGAAGAGCTCTGCGC 1119
DB	2155125	GACACAGTTCCTCAGATCTCTGACGCTAGGCTGGAGCTGACGAAGAGCCCTGCT 2155066
QY	1120	GTTGCCCAACCTTGGGAACCGATGATACATCAAAAGGAAGAAACGCGGTGACGCGCAGC 1179
DB	2155065	GCGGCTTGGCAATACACCGCAAGGAGCGCGCGGTGCGCGACCATCCCGGACCGCTTC 2155006
QY	1180	GACCCCTCTATTGGCCACAAACCGGATCATGACCGATGCGGATATGCGGATTAAGGTAAT 1239

DB	2155005	GGGGGCC---AGGGCGCTCCCGACGATGCTGGCCACTGACCTCTCGCTGCGGTTGGAT 2154949
QY	1240	CCGACCTATCGCTATCTCGGAATAATTCATGGCCGATCCTGAGTACTTCAAGAAACT 1299
DB	2154948	CCGATTTATGAGGGATACGCGTGGTGGTGGAAACACCCGAGGAATTTGGCGACGAG 2154889
QY	1300	TTCCGAAGCGTGGTTCAAGCTCAGCACCGCTGACCTGGCCCGGAAATCACGTTTACATC 1359
DB	2154888	TTCCGAAGCGCTGGTACAGCTGATCCACGAGACATGGTCCGTTGGAGATACCTT 2154829
QY	1360	GGCCCGGAAGTGGCGGAGAGACCTGATTTGGCAAGACCCGATTCGGGAGGTAAACAC 1419
DB	2154828	GGCCCGCTGGTCCCAAGCAGACCTGCTGCTGGCAGGATCCGGTCCCTGGGTCAGCCAC 2154769
QY	1420	GACTACTGCGAAGAAGTGG-----TCAAGCAGAAATTCACAAAGTGGCCTG 1467
DB	2154768	GACTGCTGGCGAAGCGGAGATTCAGAGCCGATTCGGGGATTCGCGGATG 2154709
QY	1468	AGCATTAGTGATGATGGTCTCACCGCTTGGGACAGTGCCTGATCTTATCCGGTTCGAT 1527
DB	2154708	ACTGTCTCACAGCTAGTTTCGACCGCATGGCGCGCGCTGCTGCTCCGTTGGTAGCGAC 2154649
QY	1528	ATCGGGCGGCTGCTAACGGTGGCCGATTTCCGCTTGGCCCGCCACAGAACAGTGGCAGGGC 1587
DB	2154648	AAGCGCGCGCGCCCAACGGTGGTGGCTCCGCTGCAGCCACAAAGTCCGGTGGGAGGTC 2154589
QY	1588	AACGAGCCGA---GGCCCTGGCGAAGTCTGAGCGCTCTACGAGCAGATCT----- 1636
DB	2154588	AACGACCCCGCGGAGTTCGCCAAGGTCAITTCGCCACCTTGAAGAGATCCAGGAGTCA 2154529
QY	1637	-----CTGCCGACACCGCGCTAGCATCGCGACGCTGATCGCTTCGTGCGC 1680
DB	2154528	TTCAACTCCGCGCGCGGGAACATCAAGTGTCTTCCCGACCTCGTCTGCTCGGT 2154469
QY	1681	GGTAGGCTAGGCAAGCGCGGAAAGCAGCAGAGTTACGATGTCGCGTTCCTTC 1740
DB	2154468	GGCTGTGCCCCATAGAAAGCAGCAAGCGCGCTGGCCACCAACATCACGGTCCCTTC 2154409
QY	1741	CTGAAAGCGCTGGCGATGCGCGCGAGATGACCGACGACAGCTCTTCGACACCGCTG 1800
DB	2154408	ACCCCGCGCGCACGGATGGCTGCCAGGAACAACCGACCTGGAATCTTTGCGGTGCTG 2154349
QY	1801	GAGCGGTGCGCATGGCTTCCGAACTGGCAGAAAGAGATGATGTGTGAAGCCGAA 1860
DB	2154348	GAGCCCAAGCAGATGGCTTCCGAACTACCTCGGAAAGGCAACCGTTTGGCGCGGAG 2154289
QY	1861	GAGATGCTGCTGATGCTGGCAGCTGATGGGCTTAAACCGCCCGGAAATGACCGTGTG 1920
DB	2154288	TACATGCTGCTGACAAAGCGCAACCTGCTTACGCTCAGTGGCCCTGAGATGACGGTGTG 2154229
QY	1921	CTGGCGGTATGCGGCTACTGGGCACCACTATGGTGGCAACCAACAGCGCTATTACAC 1980
DB	2154228	GTAGGTGGCTGGCGCTCTCGCGGCAAACTACAGGCTTACCGCTGGCGGTGTTACAC 2154169
QY	1981	GATTGTGAAGCCGATGACCAACGACTTTTTTGTGAACCTGACCGATATGGGAAACAGC 2040
DB	2154168	GAGGCTCCGAGTCACTGACCAAGACTTCTTGTGAACCTGCTCGACATGGGTATCAC 2154109
QY	2041	TGGAAGC---CGTAGGTAGCAACGCTTACGAAATTCGGGACCGCAAGACCGGTGGCTG 2097
DB	2154108	TGGGAGCCCTCGCCAGCAGATGACGGGACCTACAGGGCAAGGATGCGAGTGCAGAGGTG 2154049
QY	2098	AGCTGACCGCTTCGCGGTGGATGCTGTTTGGTTCGAACCTGCTACTGCGCTCTTAC 2157
DB	2154048	AAGTGACCGCGACCGCTGGGACCTGGTCTTCGGGTTCGAACCTCGGAGTTGGGGGCTT 2153989
QY	2158	GCAGAGTGTACCCCGAGGACGATTAACGGCGAGAAAGTTTCCAACTGCTACTGCGCTTAC 2217
DB	2153988	GTGAGGTCTATGGCGCGGATGACGCGACCGCAAGTTCGTGAGGAGCTTCGTCGCTGCC 2153929
QY	2218	TGACCAAAAGTGAAGACCGCCGACCGTTTCGAGCT 2252

QY 638 AAAAGATATCTACTGGGTGACGAAAGAGAGTGGCTGGCACCTTCTGTGACGAAAGCTACG 697
 Db 590 ---ATGAGGTCTATTGGGGCAAGAAAGCCACTGGCT-----CGGCG 628
 QY 698 CGCAGCTGAACAAGCCAGAGACCAATGGAATAACCCGCTGGCGGTGTCCAAATGGTGTGTA 757
 Db 629 ATGACGGTTACAGCGTAAAGCATCTGGNAGAACCCGCTGGCGCGGTGCNATGGGGCTGA 688
 QY 758 TCTATGTGAACCCGGAAGGTGTTAACGGCCACCTGTATGCCCTGAGAACCCGACAGG 817
 Db 689 TCTACGTGAACCCGAGCGCGCAACGCAACACCGGACCCCATGGCCGCGGCTGCGACA 748
 QY 818 TACTTGAACCTTCGCCGTATGCGATGACGAGCAAGAAACCGCAGCCCTCAGACTG 877
 Db 749 TTCGGAGACGTTTCGGCGCATGCGCATGAACAGCTGCGAAGCGCGCTGTATCGTCG 808
 QY 878 CGCGCCACACCGCTGGTAAATTTGTACGCTAATGGCAATGCCCTCTGCGTTAGCCCTGACC 937
 Db 809 CGGTCACACTTTCGGTAAGACCCATGGCGCGCGCCGCGGATCTGGTCGGCCCGGAC 868
 QY 938 CAAAGCCTCTGACGTTGAAACAGGGCTTTAGTTGGGGCAACCCCAACATGCAAGGCA 997
 Db 869 CCGAGGCTGCTCCGCTGAGCAGATGGCTTTGGCTGGAAGAGCTCGTATGCCACCGGAA 928
 QY 998 AGGCAAGCAACGCGTACCTCGGTATCGAGGTGCTTGACACCAACCCACGAAAT 1057
 Db 929 CCGTAAAGAGCGCATCACAGCGGCATFCGAGGTGCTATGACCAACACCCCGACGAAAT 988
 QY 1058 TCGATATGGGCTATTTCCACCTGCTGTTGGCTACAATTTGGAATGGAAGAGTCCGT 1117
 Db 989 GGGACACAGTTTCTCGAGATCTGTACGGCTACGAGTGGAGCTGACGAAGAGCCCTG 1048
 QY 1118 CCGTGGCCCACTTGGGAACCGATTTGACATCAAAAAGGAANAACAAGCCGTTGACGCCA 1177
 Db 1049 CTGGCGCTTGGCAATACACCCCAAGGACGGCGCGGTGCGCGCACCATCCCGACCCGT 1108
 QY 1178 GCGACCCCTCTATTTCGCCACAAACCGCATGACCGATGCGGATGCGGATTAAGGTAA 1237
 Db 1109 TCGGGGGGCC---AGGGGCTTCCCGACGATGCTGGCCACATGACCTCTCGCTGGGGGTG 1165
 QY 1238 ATCCGACCTATCGGCTATCTCGAATAATTCATGGCCGATCCTGAGTACTTCAAGAAA 1297
 Db 1166 ATCCGATCTATGAGCGGATACGCGTCGCTGGCTGGAACACCCCGAGGAATTTGCCGAG 1225
 QY 1298 CTTTCGCGAAGGCTGTAAAGCTGACGCAACCGTGACCTGGGCCCGCAAAATCAGCTTACA 1357
 Db 1226 AGTTCCGAAGCGCTGTACAGCTGATCCACCGAGACATGGTCCCGTTGCGAGATACC 1285
 QY 1358 TCGGCCGGAAGTCCGCGCAGACGATGATTGGCAAGACCCGATTCGGCAGGTACA 1417
 Db 1286 TTGGCCGCTGGTCCCCAAGCAGACCTGCTGTGGCAGGATCCCGTCCCTGCGGTACGA 1345
 QY 1418 C--CGACTACTCGGAAGAGTGGTCAAGCAGAAAATTCACAAAGTGGCCCTGAGCAT-- 1473
 Db 1346 CGACCTCGTCGCGNAGACAGATTGCCAGCCCTTAGAGCCAGATCCGGGCATCGGGATTGA 1405
 QY 1474 -----AGTGAATGCTTCCACCGCTTGGGACAGTGCCTGCTATATCCGCTTCCGATA 1528
 Db 1406 CTGTCTCACAGCTAGTTTCGACCGCATGGCGGGCGCTCGCTCGTTCCGTGTAGCGACA 1465
 QY 1529 TGGCGGGGTGCTTAACGTTGCCCATTCGCTTGGCCCCACAGNACGAGTGGCAGGCA 1588
 Db 1466 AGCGGGCGGCGCAACGGTGGTCATCCGCTTCACGCCACAAGTGGGTGGGAGTCA 1525
 QY 1589 ACGAGCCGAGCGCTGGCGAAAGTGCCTGAGCGTCTACGACGAGATCTCTGCGC----- 1642
 Db 1526 ACGACCCGAGGATCTCGCAAGGTCAATTGCGACCCCTGAAGAGATCCAGGATCATTCA 1585
 QY 1643 -----ACACCGCGCTAGCATTCGCGGAGTGATGCTTCTGGCCGCTAGCGTAG 1690
 Db 1586 CTCGGCGGGGAACATCAAAAGTCTCTTCGCCGACCTCGTCTGCTCGTGGCTGTGCGC 1645
 QY 1691 GCATCGAGAAAGCCGCAAGCAGCAGGTAGCATGTGCGGGTTCCTCTCTGAAGGCC 1750

Db 1646 CACTAGAGAAAGCAGAAAGGGGCTGGCCAAACATCACGGTGGCCCTTCACCCCGGGCC 1705
 QY 1751 GTGGCGATCGCACGCCCGAGATGACCGACGCACTCTCTTGCACCGCTGGAGCGCTGG 1810
 Db 1706 CGCAGATGCGTCGCGAGGAACAAACCGAGCTGGAATCTTTGCCGTGTGGNGCCCAAG 1765
 QY 1811 CCGATGGCTTCGCAACTGGCAGAAAGAGTATGTGGTGAAGCCGGAAGATGCTGTC 1870
 Db 1766 CAGATGGCTTCGAAACTACCTCGAAAGGCAACCGTTGCGGCGGAGTACATCGCTGC 1825
 QY 1871 TGGATCGTGGCAGCTGATGGCTTAACGGCCCCGGAATGACCCGTGCTGTGGCGGTA 1930
 Db 1826 TCGACAAGCGCAACCTGCTTAGCTCAGTGGCCCTGAGATGACGGTGTGGTAGTGGCC 1885
 QY 1931 TCCGCTACTGGCACCACCAACTATGTGGCACCACCAACCGGCTATTTCACCGATTTCTGAAG 1990
 Db 1886 TGGCGTCTCGGCGCAACTACAGCGCTTACCGCTGGCGGTTCACCGAGGCGCTCG 1945
 QY 1991 GCCAGTTGACCAAGACTTTTTTTGTGAACCTGACCGATATGGGGAACAGCTGGAAGC--- 2047
 Db 1946 AGTCACTACCAACGACTTCTTCGTGAACCTGCTCGACATGGGTATCACCTGGGAGCCT 2005
 QY 2048 CGGTAGTAGCAACGCCCTACGAAATCCCGGACCGCAAGACCGGTGCGGTGAAGTGGACG 2107
 Db 2006 CGCCAGCAGATGACGGGACCTTACCAGGGCAAGGATGGCAGTGGCAAGGTGAAGTGGACG 2065
 QY 2108 CTGCGCGGTGATCTGGTATTGTTCCAACTCCTACTGCGCTTACGCGCAGAGTGT 2167
 Db 2056 GCAGCCGCTGGACCTGGTCTTCGGGTCCAACTCGGAGTTCGGCGGCTTTCGAGGTCT 2125
 QY 2168 ACGCCAGGACGATTAACGGCGAGAAAGTTCGTAGAGACTTCGTCCGCCCTGGACCAAG 2227
 Db 2126 ATGCGCGATACGCGCGCAGCGAAAGTTTCGTACAGGATTCGTCTGCTGGGACAAG 2185
 QY 2228 TGATCAACCGCGACCGTTTCGACGT 2252
 Db 2186 TGATGAACCTGACAGGTTCGACGT 2210

RESULT 15
 AAQ51531
 ID AAQ51531 standard; DNA; 4795 BP.
 XX AC AAQ51531;
 XX DT 25-MAR-2003 (updated)
 XX DT 17-MAY-1994 (first entry)
 XX DE M.tuberculosis H37Rv katG gene.
 KW catalase-peroxidase; isonicotinic acid hydrazide; isoniazid; INH;
 KW antibiotic; susceptibility; sensitive; resistant; katG; ss.
 XX OS Mycobacterium tuberculosis (strain H37Rv).
 XX Key Location/Qualifiers
 XX CDS 179..4186
 XX /tag= a
 XX /product= catalase-peroxidase
 XX
 XX WO9322454-A1.
 XX 11-NOV-1993.
 XX 30-APR-1993; 93WO-EP01063.
 XX 17-SEP-1992; 92FR-0011098.
 XX 30-APR-1992; 92US-0875940.
 XX 14-AUG-1992; 92US-0929206.
 XX 16-APR-1993; 93FR-0004545.
 XX (ASSI-) ASSISTANCE PUBLIQUE.
 PA

PA (INSP) INST PASTEUR.
 PA (MEDI-) MEDICAL RES COUNCIL.
 PA (UYBE-) UNIV BERNE.
 PA (UYPA-) UNIV CURIE PARIS VI P & M.
 XX Bodmer T, Cole S, Heym B, Honore N, Telenti A;
 PI Young D, Zhang Y;
 XX WPI: 1993-368812/46.
 DR P-PSDB; AAR43670.
 XX
 PT Rapid detection of antibiotic resistance in Mycobacteria - esp.
 PT Isoniazid, rifampicin or streptomycin resistance in tuberculosis
 PT by detecting mutation in katG, rpoB or rpsL genes
 XX
 PS Example 2; Fig 6A; 97pp; English.
 XX
 CC The katG gene of Mycobacterium tuberculosis was isolated on a KpnI
 CC fragment by shotgun cloning. The sequence contains one open reading
 CC frame with high coding probability; a 735 amino acid protein is
 CC encoded by the ORF having predicted mol.wt. 80029. This protein
 CC conforms to the structure of known bacterial catalase-peroxidases
 CC which consist of two modules (possibly the result of a duplication
 CC event), both showing homology to the yeast enzyme, fused to a unique
 CC N-terminal sequence of about 50 amino acid residues. Mutations in
 CC the katG gene can result in loss of INH-susceptibility, i.e.
 CC the formation of antibiotic resistant Mycobacterial strains
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 4795 BP; 892 A; 1532 C; 1559 G; 812 T; 0 other;
 Query Match 25.6%; Score 578.2; DB 14; Length 4795;
 Best Local Similarity 57.8%; Pred. No. 1.5e-153;
 Matches 1229; Conservative 0; Mismatches 838; Indels 58; Gaps 9;
 QY 160 AACAAAGATGGTGGCGGAGGTTGAACCTCGATATTTTGCATAGCAAGATCGGAAA 219
 DB 160 AAA 219
 QY 2081 AACAGAGCTGGTGGCGGAGGTTGAACCTCGATATTTTGCATAGCAAGATCGGAAA 210
 DB 2081 AAA 210
 QY 220 TCAGACCCGATGGTGGCGGAGGTTGAACCTCGATATTTTGCATAGCAAGATCGGAAA 279
 DB 220 AAA 279
 QY 2141 GCTGACCCGATGGTGGCGGAGGTTGAACCTCGATATTTTGCATAGCAAGATCGGAAA 2200
 DB 2141 AAA 2200
 QY 280 GCCTGAAGAAAGATGTCCACCGGTTGATGACCGATAGCAAGATGGTGGCGGAGTAC 339
 DB 280 AAA 339
 QY 2201 GCCTGACGCGGAGATCGAGGAAGTATGACCACTCGGAGCGGTGGTGGCGGCGGAC 2260
 DB 2201 AAA 2260
 QY 340 TGGGGCACTACGCGGTTGATGATCCGTTATGGCTTGGCACTCCGCTGGCACTACCGT 399
 DB 340 AAA 399
 QY 2261 TAGCGCACTACGCGGCGGTGTTATCCGGATGGTGGCACTCCGCTGGCACTACCGT 2320
 DB 2261 AAA 2320
 QY 400 ATTGCTGATGGCGGTGGCGGAGGTTGATGACCGGAGGAGGTTGACCGCTCAACTCC 459
 DB 400 AAA 459
 QY 2321 ATCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTAACAGC 2380
 DB 2321 AAA 2380
 QY 460 TGGCGGAGCACTGACCGTTGATGAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGTCA 519
 DB 460 AAA 519
 QY 2381 TGGCGGAGCACTGACCGTTGATGAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGTCA 2440
 DB 2381 AAA 2440
 QY 520 TAGGGCAACAAATACGCTGGCGAGACCTGATGATTTCTGGCTGGCA--CCGTGGCTATG 577
 DB 520 AAA 577
 QY 2441 TAGGGCAACAAATACGCTGGCGAGGAGCTGATTTTTCGGCGGAGCGCGCTGGCTCGG 2500
 DB 2441 AAA 2500
 QY 578 AGTCCATGGCTTACCTGCTTACGCTTCTTCTTGGCGCGCGCGCGCGCGCGCGCGCG 637
 DB 578 AAA 637
 QY 2501 AATCGATGGCTTACAGAGTTTCGGGTTTCGGTTTCGG--CGCTCGACGAGTGGGAGCCG 2558
 DB 2501 AAA 2558
 QY 638 AAAAGATATCTACTGGGTGAGCAAAAAGAGTGGCTGGCACCTTCTGACGAACGCTACG 697
 DB 638 AAA 697
 QY 2559 ---ATGAGTCTATTGGGCAAGGAAGCAACCTGGCT-----CGGCG 2597
 DB 2559 AAA 2597
 QY 698 GCGACGTGAACAGCCAGAGACCATGGAACCCGCTGGCGGTGTCGAATGGGTCTGA 757
 DB 698 AAA 757

DB 2598 ATGACGGTTACAGCTAAGCGATCTGAGAAACCCGCTGGCGGCTGACAGTGGGCTGA 2657
 QY 758 TCTATGTGAACCGGAGGTTTAAACGGCCACCTGATCCGCTGAGAAACCGCAGCAGG 817
 DB 2658 TCTACGTGAACCGGAGGCGGCAACCGGACCCGATGCGCGCGGCGGCTGACGA 2717
 QY 818 TACTTGAACCTTTCGCCGCTATGCGGATGAACGAGAAACCGGACCGCTCACAGCTG 877
 DB 2718 TTGCGGAGAGCTTTCGCCGCTATGCGGATGAACGAGCTGGAACAGCGCGCTGATCTG 2777
 QY 878 GCGGCCACACCGCTGCGTAAATGTCAGGTAATGGAATGCTCTGCTTGGCTTACGCCCTGACC 937
 DB 2778 GCGGTACACTTTCGTTAAGACCATGGCGCGCGCGCGGATCTGCTGGCGCGCAAC 2837
 QY 938 CAAAGCTCTGACGTTGAACACGAGGCTTAGTGGGCGCAACCCCAACATCGAGGCA 997
 DB 2838 CCGAGGCTGCTCGCGCTGGAGCAGATGGCTTGGCTGGAAGAGCTGATGGAACCGGAA 2897
 QY 998 AGCAAGCAACCGCGCTGACCTCGGGTATCGAAGGTGCTTGGACCAACCCCAACGAAAT 1057
 DB 2898 CCGTAAAGGACGCGATCACCGAGGCGATCGAGGTGCTGATGAGCAACACCCCGAGAAAT 2957
 QY 1058 TCGATATGGCTTATTCGACCTGCTGTTGGCTACATTTGGGACTGAAAGAGTCTCTG 1117
 DB 2958 GGGACAAACAGTTTCTCGAGATCTGTCAGGCTACGAGTGGGAGCTGACGAAGAGCCCTG 3017
 QY 1118 CCGGTGCCACCATTTGGGAACCGATGACATCAAAAGAGAAACCAACCGGTTGACGCCA 1177
 DB 3018 CTGGCGTTGGCAATACACCGCCAGGCGCGCGGCTGGCGGACCATCCCGGACCCGT 3077
 QY 1178 GCGACCCCTCTATTCGCGCACACCGCATCATACCGATGCGGATATGGCGATAAAGTAA 1237
 DB 3078 TCGCGCGGCGC---AGGCGCGCTCCCGAGGATGCTGGCCACTGACCTCTCGCTCGGGTGG 3134
 QY 1238 ATCGACCTATCGCGCTATCTGCGGAAATTTCTAGCCGCTGCTGAGTACTTCAAGAAAA 1297
 DB 3135 ATTCGATCTATGACCGGATCACGCTGCTGGTGGAAACACCCGAGGAAATGGCCGAGC 3194
 QY 1298 CTTTTCGGAAGGCTGGTTCAAGCTGACGACCGCTGACCTGGCGCGCGGAAATACGTTTACA 1357
 DB 3195 AGTTTCGCAAGGCTGGTACAGCTGATCCACGAGCATGGTCCGCTGCGGATGAC 3254
 QY 1358 TCGCGCGGAAAGTTCGCGGCAAGAGACTGATTTTGGCAAGACCGGATTTCCGCGAGTAAACA 1417
 DB 3255 TTGGCGCGCTGGTCCCAAGCAGACCGCTGCTGGCAGGATCCGCTCCCTCGCTGACGA 3314
 QY 1418 C--CGACTACTCGGAAGAGTGTCAAGCAGAAATTTGCAAAAGTGGCTGAGCATT-- 1473
 DB 3315 CGACCTCGTGGCGGAGCAGATTGGCAGCCCTTAAGAGCCAGATCCGCGGATTTGA 3374
 QY 1474 -----AGTGAGATGGTCTCCACCGCTTGGGACAGTCCCGTACTTATCGCGGTTCCGATA 1528
 DB 3375 CTGCTCACAGCTAGTTTCGACCGCATGGCGGCGGCTGCTGTTCCGTTGGTAGCAGCA 3434
 QY 1529 TGGCGGCGGTGTAAAGTGGCGGATTCGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCA 1588
 DB 3435 AGCGCGCGGCGGCAACCGTGGTGGCTCGCATCCGCTGACGACCAAGTGGTGGGAGGTCA 3494
 QY 1589 ACCGAGCGGAGCGCTGGCGGAAAGTGTGAGCGCTTACGAGCAGATCTTGGCGG----- 1642
 DB 3495 ACACCGGCGGAGCTGTCGCGCAAGGCTATTCGACCCCTGGAAGAGATCCAGGAGTCAATCA 3554
 QY 1643 -----ACACCGGCGCTAGCATCGGACGCTGATCGTTCTTGGCGGCTAGCGTAG 1690
 DB 3555 CTCGCGCGGGAACATCAAGTGTCTTCCGCGACCTGCTGCTGCTGCTGCTGCTGCTGCTG 3614
 QY 1691 GCATCGAGAAACCGGAAAGCAGAGGTTACGATGTGCGCGCTTCCCTTCCCTGAAAGGCC 1750
 DB 3615 CACTAGAGAAAGCAGAAAGGCGGCTGGCCACCAACATACGCGTGGCTTCCACCCCGGCGC 3674
 QY 1751 GTGGCATGGCGCGCGGAGATGCCGACGACACTCTTCGACCGCGCTGGAGCGCGCTGG 1810
 DB 3675 CGCACGATGGTGGCGAGGAAACACCGAGCTGGAATCTCTTGGCTGCTGGAGCCCAAGG 3734

THIS PAGE BLANK (USPTO)

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	2262	100.0	2262	10	US-09-684-889-5		Sequence 5, Appli
2	672.2	29.7	2238	10	US-09-684-889-7		Sequence 7, Appli
c 3	618.6	27.3	82993	15	US-10-080-170-645		Sequence 645, App
c 4	137.2	6.1	620	11	US-09-940-925A-144		Sequence 144, App
5	137.2	6.1	620	11	US-09-940-925A-146		Sequence 146, App
c 6	137.2	6.1	620	11	US-09-940-925A-148		Sequence 148, App
c 7	137.2	6.1	620	11	US-09-940-925A-150		Sequence 150, App
8	137.2	6.1	620	11	US-09-941-193A-144		Sequence 144, App
9	137.2	6.1	620	11	US-09-941-193A-146		Sequence 146, App
c 10	137.2	6.1	620	11	US-09-941-193A-148		Sequence 148, App
c 11	137.2	6.1	620	11	US-09-941-193A-150		Sequence 150, App
12	135.6	6.0	620	11	US-09-940-925A-143		Sequence 143, App
c 13	135.6	6.0	620	11	US-09-940-925A-145		Sequence 145, App
c 14	135.6	6.0	620	11	US-09-940-925A-147		Sequence 147, App
c 15	135.6	6.0	620	11	US-09-940-925A-149		Sequence 149, App
16	135.6	6.0	620	11	US-09-941-193A-143		Sequence 143, App

QY 121 CCTGTCATGACGGTGGTAACACCTCGACCGGTACTTCCAAACAAGATTGGTGGCCGAA 180
DB 121 CCTGTCATGACGGTGGTAACACCTCGACCGGTACTTCCAAACAAGATTGGTGGCCGAA 180
QY 181 GGGTTGAACCTGGATATTTTGCATCAGCAAGATCGCAATCAGACCCGATGGATCCGGAT 240
DB 181 GGGTTGAACCTGGATATTTTGCATCAGCAAGATCGCAATCAGACCCGATGGATCCGGAT 240
QY 241 TTCAACTACCGTGAAGAAGTACGCAAGCTCGATTTCGACGCGTGAAGAAGATGTCCAC 300
DB 241 TTCAACTACCGTGAAGAAGTACGCAAGCTCGATTTCGACGCGTGAAGAAGATGTCCAC 300
QY 301 CGGTTGATACCGATACCAAGAGTGTGGCCCGCTGACTGGGGCACTACGGGGGTTG 360
DB 301 CGGTTGATACCGATACCAAGAGTGTGGCCCGCTGACTGGGGCACTACGGGGGTTG 360
QY 361 ATGATCCGTTATGGCTTGGCACTCCGCTGGCACCTTACCGTATGCTGATGGCGGGGC 420
DB 361 ATGATCCGTTATGGCTTGGCACTCCGCTGGCACCTTACCGTATGCTGATGGCGGGGC 420
QY 421 GGTGGTACCGGAAGCCAGCGCTTTGACCGCTCAACTCCTGGCCGGAACAAGTACGCTG 480
DB 421 GGTGGTACCGGAAGCCAGCGCTTTGACCGCTCAACTCCTGGCCGGAACAAGTACGCTG 480
QY 481 GATAAAGCGCGCTCTGCTGGCCGATCAAGAGAAGTACGCAACAATACTAGCTGG 540
DB 481 GATAAAGCGCGCTCTGCTGGCCGATCAAGAGAAGTACGCAACAATACTAGCTGG 540
QY 541 GCAGACCTGATGATTCTGGCTGGCACCGTGGCTTATGAGTCCATGGGCTTACCTGCTTAC 600
DB 541 GCAGACCTGATGATTCTGGCTGGCACCGTGGCTTATGAGTCCATGGGCTTACCTGCTTAC 600
QY 601 GGCTTCTTTTGGCCCGCTCGATATTTGGGAACCCGGAAGATATCTACTGGGGTGAC 660
DB 601 GGCTTCTTTTGGCCCGCTCGATATTTGGGAACCCGGAAGATATCTACTGGGGTGAC 660
QY 661 GAAAAACCGCTGGCGGCTGTCCAAATGGGTCTGATCTATGTGAACCCGGGAAGTGT 780
DB 661 GAAAAACCGCTGGCGGCTGTCCAAATGGGTCTGATCTATGTGAACCCGGGAAGTGT 780
QY 781 AACGGCCACCTGATTCGGCTAGAACCGCACAGAGTACTTGAACCTTTCGGCCGATG 840
DB 781 AACGGCCACCTGATTCGGCTAGAACCGCACAGAGTACTTGAACCTTTCGGCCGATG 840
QY 841 GCGATGAACGACGAAAAAACCGGACCGCTCACAGCTGGCGGCCACACCGTGGTAATTGT 900
DB 841 GCGATGAACGACGAAAAAACCGGACCGCTCACAGCTGGCGGCCACACCGTGGTAATTGT 900
QY 901 CACGGTAAATGGCAATGCTTGGCTTAGCCCTGACCCAAAGCCTCTGACGTTGAAAC 960
DB 901 CACGGTAAATGGCAATGCTTGGCTTAGCCCTGACCCAAAGCCTCTGACGTTGAAAC 960
QY 961 CAGGGCTTAGTTGGGCAACCCCAACATGAGGGCAAGGCAAGCCGCTGACCTG 1020
DB 961 CAGGGCTTAGTTGGGCAACCCCAACATGAGGGCAAGGCAAGCCGCTGACCTG 1020
QY 1021 GGTATCAAGTGTCTTGGACCAACCCCAAGAAATTCGATATGGGCTATTTTCGACCTG 1080
DB 1021 GGTATCAAGTGTCTTGGACCAACCCCAAGAAATTCGATATGGGCTATTTTCGACCTG 1080
QY 1081 CTGTTCCGCTACAAATTTGGGAACGAAAAAGAGTCTTCCCGGTCGCCACCTATTTGGGACCG 1140
DB 1081 CTGTTCCGCTACAAATTTGGGAACGAAAAAGAGTCTTCCCGGTCGCCACCTATTTGGGACCG 1140
QY 1141 ATTGACATCAAAAGGAAACAGCGGGTTGACGCCAGGACCGCTTATTCGCCAAC 1200
DB 1141 ATTGACATCAAAAGGAAACAGCGGGTTGACGCCAGGACCGCTTATTCGCCAAC 1200
QY 1201 CCGATCATGACCGATGGCGATATGGCGATAAAGTAAATCCGACCTATCCGCTATCTGCG 1260

DB 1201 CCGATCATGACCGATGGCGATATGGCGATAAAGTAAATCCGACCTATCCGCTATCTGCG 1260
QY 1361 GAAAAATTCATGGCCGATCTCTGAGTACTTCAAGAAAATCTTCGGAAGGCGTGGTTCAAG 1320
DB 1361 GAAAAATTCATGGCCGATCTCTGAGTACTTCAAGAAAATCTTCGGAAGGCGTGGTTCAAG 1320
QY 1321 CTGACGACCGTGCACCTGGGCCCGAAATACGTTTACATCGGCCCGGAAGTTCGCCGAGAA 1380
DB 1321 CTGACGACCGTGCACCTGGGCCCGAAATACGTTTACATCGGCCCGGAAGTTCGCCGAGAA 1380
QY 1381 GACCTGATTTGGCAAGACCCGATTCGCGCAGGTAAACACGACTACTCGGAAGAAGTGGTC 1440
DB 1381 GACCTGATTTGGCAAGACCCGATTCGCGCAGGTAAACACGACTACTCGGAAGAAGTGGTC 1440
QY 1441 AAGCAGAAAAATTCGCAAAAGTGGCTGAGCATTTAGTGTCTCTCCACCGCTTGGGAC 1500
DB 1441 AAGCAGAAAAATTCGCAAAAGTGGCTGAGCATTTAGTGTCTCTCCACCGCTTGGGAC 1500
QY 1501 AGTGCCCGTACTTATCGCGGTTCCGATATGCGCGCGGTGCTAACGGTGCCCGCATTTCCG 1560
DB 1501 AGTGCCCGTACTTATCGCGGTTCCGATATGCGCGCGGTGCTAACGGTGCCCGCATTTCCG 1560
QY 1561 TTGGCCCGCACAGACGAGTGGCAGGGCAACGAGCGCGAGCCCTGGCGAAAGTCTGAGC 1620
DB 1561 TTGGCCCGCACAGACGAGTGGCAGGGCAACGAGCGCGAGCCCTGGCGAAAGTCTGAGC 1620
QY 1621 GTCTACGAGCAGATCTCTGCGCACCGCGCTAGCATTCGCGGACGTGATGCTGCTGGCC 1680
DB 1621 GTCTACGAGCAGATCTCTGCGCACCGCGCTAGCATTCGCGGACGTGATGCTGCTGGCC 1680
QY 1681 GGTAGCTAGGCAATCGAGAAAGCCGGAAGACGAGTTACGATGTCGCGTTCCCTTC 1740
DB 1681 GGTAGCTAGGCAATCGAGAAAGCCGGAAGACGAGTTACGATGTCGCGTTCCCTTC 1740
QY 1741 CTGAAAGCCGTTGGCGATGCGACCGCGAGATGACCCAGCAGACTCCTTCGCACCGCTG 1800
DB 1741 CTGAAAGCCGTTGGCGATGCGACCGCGAGATGACCCAGCAGACTCCTTCGCACCGCTG 1800
QY 1801 GAGCGCTGCGCGATGGCTTCCGCACTGCGCAAGAAAGATGATGTTGGAAGCCGGA 1860
DB 1801 GAGCGCTGCGCGATGGCTTCCGCACTGCGCAAGAAAGATGATGTTGGAAGCCGGA 1860
QY 1861 GAGATGCTGCTGGATCTGCGCAGCTGATGGGCTTAAACCGCCCGGAAATGACCGTCTG 1920
DB 1861 GAGATGCTGCTGGATCTGCGCAGCTGATGGGCTTAAACCGCCCGGAAATGACCGTCTG 1920
QY 1921 CTGGCGGATGCGCGTACTGGGACCAACTATGGTGGCAACACGCGGCTATTTCACC 1980
DB 1921 CTGGCGGATGCGCGTACTGGGACCAACTATGGTGGCAACACGCGGCTATTTCACC 1980
QY 1981 GATTGTGAAGCCAGTTGACCAACGACTTTTGTGAACCTGACCGATATGGGGAACAGC 2040
DB 1981 GATTGTGAAGCCAGTTGACCAACGACTTTTGTGAACCTGACCGATATGGGGAACAGC 2040
QY 2041 TGAAGCCGCTAGTAGCAACGCGTACGAAATCCGCAACCGCAGACCGCGTGGTGAAG 2100
DB 2041 TGAAGCCGCTAGTAGCAACGCGTACGAAATCCGCAACCGCAGACCGCGTGGTGAAG 2100
QY 2101 TGAACCCCTCGCGGTTGGATCTGGTATTTGGTTTCCAACCTCGCTACTGCGCTCTTAGCA 2160
DB 2101 TGAACCCCTCGCGGTTGGATCTGGTATTTGGTTTCCAACCTCGCTACTGCGCTCTTAGCA 2160
QY 2161 GAAGTGTACGCCAGGACGATAACGGGAGAAGTTCTCAGAGACTTTCGTCGCGCGCTGG 2220
DB 2161 GAAGTGTACGCCAGGACGATAACGGGAGAAGTTCTCAGAGACTTTCGTCGCGCGCTGG 2220
QY 2221 ACCAAAGTATGAACGCCGACCGTTTCGACGCTGCGGTCGTA 2262
DB 2221 ACCAAAGTATGAACGCCGACCGTTTCGACGCTGCGGTCGTA 2262

US-09-884-889-7
; Sequence 7, Application US/09884889
; Patent No. US20020102680A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: ROBERTSON, Dan
; APPLICANT: SANYAL, Indrajit
; APPLICANT: ADHIKARI, Robert
; TITLE OF INVENTION: CATALASES
; FILE REFERENCE: DIVER1100-4
; CURRENT APPLICATION NUMBER: US/09/884, 889
; PRIORITY FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 09/412,347
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: US 08/951,844
; PRIOR FILING DATE: 1997-10-16
; PRIOR APPLICATION NUMBER: US 08/674,887
; PRIOR FILING DATE: 1996-07-03
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 2238
; TYPE: DNA
; ORGANISM: Microscilla furvescens
US-09-884-889-7

Query Match 29.7%; Score 672.2; DB 10; Length 2238;
Best Local Similarity 59.5%; Pred. No. 1.3e-198;
Matches 1277; Conservative 0; Mismatches 808; Indels 60; Gaps 6;

QY	150	CGGTACTTCCAAACAAGATTGGTGGCCGGAAGGTTGAACCTGGATATTTTGCATCAGCA	209
DB	99	CGGCACCAAAAACAGGAGTGGTGGCCCAACATGCTCAACCTCGGCATCTTACGCCAACA	158
QY	210	AGATCGCAATCAGACCGATGGATCCGGATTCAACTACCGTGAAGAAGTACGCAAGCT	269
DB	159	TTATCGCTATCGGACCAACAGCCCGGATTTGACTATGCCGAAGATTTAAGAAGCT	218
QY	270	CGATTTCCAGCGGTGAAGAAGATGTCCAGCGTTGTATGACCGATAGCCAGATAGCAAGTGTG	329
DB	219	AGATGTGCGAGCGTTAAAGAGGACCTGGCAGCGCTAATGACAGATTACAGAGACTGGTG	278
QY	330	GCCCGCTGACTGGGGGCACTACGCGGTTTGATGATCGTATGGCTTGGCACTCCGCTGG	389
DB	279	GCCAGCAGATTACGGTCATTATGGCCCTCTTTATACGCATGGCGTGGCAGCGCCGG	338
QY	390	CACCTACCGTATTGCTGATGGCGGTGGGGGGTGGTACCGGAAGCCAGCGCTTTGCACC	449
DB	339	CACCTACCGTATCGGTGATGGCGGTGGTGGCGGTGGCTCCGCTCACAGCGCTTCGCGC	398
QY	450	GCTCACTCCTGGCGGACACAGTCAAGCTGATGATGAAGCGCGCTGCTGCTGTGGCGGAT	509
DB	399	TCTCAATAGCTGGCGAGACAATGCCAATCTGGATAAAGCAGCCTTGTCTTTGGCCCAT	458
QY	510	CAAGAAGAGTACGCAACAATCAGTGGCAGACCTGATGATCTGGCTGGGACACCGT	569
DB	459	CAACAACAATACGCTGAAATAATCTCTGGCGGATCTAATGATACACAGGAACGT	518
QY	570	GGCTTATAGTCCATGGGCTTACCTGCTTACGGCTTCTTTTCGGCGCGCTCGATATTG	629
DB	519	AGCTCTGGAACATATGGGCTTTAAACTTTTGGTTTTGCAGTGGCAGACAGATGATG	578
QY	630	GGAACCGGAAAGATATCTACTGGGTGACGAAAGAGTGGCTGGCACCTTCTGACGA	689
DB	579	GGAGCGTGAAGAAGATGATACTGGGAGCAGAACCCGAAATGGCTGGGA-----GACAA	632
QY	690	ACGCTACGCGACGTGAACAAGCCAGACCATGGAACACCGCTGGCGGCTGTCCAAAT	749
DB	633	CGGCTATGAAGGTGACCGAGAGC-----TCGAAATCCCTTGGAGCGGTACAAAT	683
QY	750	GGGTCTGATGTATGAAAGAGGTGTTAAACGGCCACCTGTATCCGCTGAGAACCGC	809
DB	684	GGGACTCATCTATGTAACCCCAAGAGCCCAAGCCAGACCATCTATCGCTGCTGC	743

QY	810	ACAGCAGGTACTTGAAACCTTTCGCCCTGTATGGCGATGAACGACGAAAAAACCGAGCCCT	869
DB	744	CGGTGATATTGCTGAGACTTTTGGCCGAATGCAATGAATGACGAAGAACCGTGGCTCT	803
QY	870	CACAGCTGGCGGCACACCGTCCGTTAAATGTCACGGTAATGGCAATGCTCTGCGT---T	926
DB	804	CATAGCGGTGGACACACCTTCGGAAAAAACCCATTCGCTGCCGATGGGAATAATGT	863
QY	927	AGCCCTGACCCAAAAGCCTCTGAGCTTGAAAACACAGGGCTTAGGTTGGGCAACCCCAA	986
DB	864	GGCCGAGAGCTTGCCTCCGCGCAGGTATTGGAAGAATGAGCCTGGGTTGAAAAACACCTA	923
QY	987	CATCAGAGGCAAGCAACGCGGTGACCTTCGGGTATCGAAGTGTCTTGACCAACCAA	1046
DB	924	CGGCACCGGACACGCTGGGATACCATCAGCTGGACTAGAAAGCGCTGGACCAAGAC	983
QY	1047	CCCCAGAAATTCGATATGGGCTATTTCGACCTGCTGTCGGCTACAAATTTGGGAACCT	1106
DB	984	CCCTACTCAATGGGACAAATACTTTTGGAAAAACCTCTTTGGTTACGAGTGGGAGCTTAC	1043
QY	1107	AAAGAGTCTCGCGGTGCCACCATTTGGGAACCGGATTGACATCAAAAAAGGAAAAACGCC	1166
DB	1044	CAAAAGTCCAGCTGGAGCTTATCAGTGGAAACCAACAGACGGTGGCGGGCTGGCACCAT	1103
QY	1167	GGTTGACGCCAGCAGCCCTCTATTTCGCCCAACCCGATCATGACCGATGCGGATATGGC	1226
DB	1104	ACCGGATGCATGATCCAGCAAGTCGACGCTCCATTTATGCTCACTAGGACCTGGC	1163
QY	1227	GATAAAGTAAATCCGACCTATCGGCTATCTGCGAAAAAATTCATGCGCGATCCCTGAGTA	1286
DB	1164	GCTGGCATGGACCTGATTACGAAAAAATTTCTCGACGGTACTATGAAAAACCTGATGA	1223
QY	1287	CTTCAAGAAAACTTTCGCGAAGCGGTGTTCAAGCTGACGACCGCTGACCTGGGCGCCGAA	1346
DB	1224	GTTCACAGATGCTTTCGCGAAGCATGTTACAACTGACACACAGATATGGGACCAA	1283
QY	1347	ATCAGCTTACATCGCCCGGGAAGTCCCGGAGAACGCTGATTTGGGAAGACCCGATGCC	1406
DB	1284	GGTGGCTACCTGGGAGCAGAGTGCCTCAGGAAGACCTCATCTGGCAAGACCTATACC	1343
QY	1407	GGCAGGTAAC-----ACCGACTACTTGGCAAGAGTGGTCAAGACAAATGGC	1454
DB	1344	AGATGTAAGCCATCTCTTTGAGACGAAAAACGATATTGAAGCCCTTAAAGCCAAATCCT	1403
QY	1455	ACAAAGTGGCTGAGCATTAGTGGTCTCCACCGCTTGGGACAGTGGCCGCTACTTA	1514
DB	1404	GGAATCGGAGCTGACGGTAAGGAGCTGGTAAGCAGCGCATGGGCTTCTGCATCTACTTT	1463
QY	1515	TGCGCGTTCCGATATGCGGCGGCTTAACGGTGCCCGCATTCGCTTGGGCCCCACAGAA	1574
DB	1464	TAGAAACTCTGACAAAGCGCGGTGCCAACGGTGCAGTATACGACTGGCCGCCACAAA	1523
QY	1575	CGAGTGGGAGGCAACGAGCGGCGCTGGCGAAGTGTGAG-----	1619
DB	1524	AGACTGGGAAGTAAACAACCTTCAGCAACTTCCAGGGTACTCAAAACACTAGAGGTAT	1583
QY	1620	-----CGTCTACGAGCAGATCTCTGCCACACACCGCGCTAGCATCGCGAGAGTAT	1670
DB	1584	CGAGGAGGACTTTAACGAGCGCAATCAGATAACAAGCAGTATCGTTGGCCGACCTGAT	1643
QY	1671	CGTTCTGGCGGTAGCGTAGGATCGAGAAACCCGGAAGCAGCAGGTTACGATGTGGG	1730
DB	1644	TGTGCTGGCGGCTGTGGGGTGTAGAAAAAGCTGCAAAAAGATGCTGGCCATGAGGTGCA	1703
QY	1731	CGTTCCCTTCTGAAAGCGGTGGCGATCGGACCGCGGAGATGACGCGCAGACTCCTTT	1790
DB	1704	GGTGCCTTCAACCGGGGACGAGCGGTGCCACCGCTGAGCAACCGGATGTGGAACTTT	1763
QY	1791	CGCACCGCTGGAGCGCTGGCGGATGGCTTCCGCAACTGGCAGAAAGAGATGTGTGT	1850
DB	1764	CGAAGCATTAGAGCCAGCGGCTGACGGCTTTTGAAGAACTACATTAACCGGAGCATAACT	1823

QY 1054 AAATTCGATATGGCTATTTCGACCTGCTGTCGGCTACATATTGGGAACCTGAAAAGAGT 1113
Db 545 AAATGGGCAACAGATTTCTCGAGATCTGTACGCTACGAGTGGGAGCTGACGAAGAGC 486
QY 1114 CCTGCGGTGCCACCAATTTGGAAACCGATTGACATCAAAAAGGAAAACAAGCCGTTGAC 1173
Db 485 CTTCTGGCGTTGGCAATACACCGCAAGGACGGCCGCTGCCGCAACCATCCCGGAC 426
QY 1174 GCCAGCAGCCCTATTTCGCCCAACACCGCATATGACCCGATGCGGATATGGCATAAAG 1233
Db 425 CCGTTCCGCGGGCC---AGGCGCTCCCGACGATGCTGGCCACTGACCTCTCGCTGGG 369
QY 1234 GTAATTCGACCTATCGCGCTATCTCGCAAAAATTCATGGCCGATCTGAGTACTTCAAG 1293
Db 368 GTGGATCGGATCTATGACGGGATCAGCGTGCCTGGTGGGAACACCCCGAGGAATGGCC 309
QY 1294 AAAACTTTCCGGAAGGGTGTTCAAGCTGACGACCGTGACCTGGGCCCCGAAATCAAGT 1353
Db 308 GACGAGTTCGCAAGGCTGTACAGCTGATCCACCGAGACATGGGTCCCGTTGCGAGA 249
QY 1354 TACATCGCCGCGGAGTGCCGGCAGAGACCTGATTTGGCAAGAAATTCGACAAAGT 1461
Db 248 AGCCACGACCTCGTCCGCGGAAGCGAGATTCACGCTTAAAGACCAAGTCCGGGCAATCG 129
QY 1462 GGCCTGAGCATTAGTGAGATGGTCTCCACCGCTTGGGACAGTGCCTGACTTATCGCGGT 1521
Db 128 GATTGACTGCTACACCTAGTTTCGACCGCATGGGCGGGCGGTCTGCTTCCGTGGT 69
QY 1522 TCCGATATGCGCGGGTGTAAACGTTGCCGATTCGCTTGGCCCCACAGAACAGATGG 1581
Db 68 AGCGACAAGCGCGCGGCGCAACGGTGTGCGCATCCGCTCGCAGCCCAAGTCGGGTGG 9
QY 1582 CAGGCCAA 1589
Db 8 GAGGTCAA 1

RESULT 7
US-09-940-925A-150/C
; Sequence 150, Application US/09940925A
; Publication No. US20030054338A1
; GENERAL INFORMATION:
; APPLICANT: BROW, MARY ANN D.
; LYAMICHEV, VICTOR I.
; OLIVE, DAVID M.
; TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
; PATHOGENS
; NUMBER OF SEQUENCES: 165
; CORRESPONDENCE ADDRESS:
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/940,925A
; FILING DATE: 10-Jun-2002
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: FORS-01756
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 150:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 620 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 150:
US-09-940-925A-150

Query Match 6.1%; Score 137.2; DB 11; Length 620;
Best Local Similarity 54.3%; Pred. No. 4.7e-32;
Matches 330; Conservative 0; Mismatches 263; Indels 15; Gaps 2;

QY 994 GCGAAGGCAACACGCGTGACCTCGGGTATCGAAGGTGTTGGACCAACCAACCCACG 1053
Db 605 GGAACCGGTAGGACGCGATCACCAACGCGCATCGAGTCTGATGACCAACACCCGAGC 546
QY 1054 AAATTCGATATGGGCTATTTCGACCTGCTTTCGGCTACAATTTGGGAACCTGAAAAGAGT 1113
Db 545 AAATGGGCAACAGATTTCTCGAGATCTGTACGGCTACGAGTGGGAGCTGACGAAGAGC 486
QY 1114 CTTGCCGTTGCCACCATTTGGAAACCGATTGACATCAAAAAGGAAAACAAGCCGTTGAC 1173
Db 485 CTTGCTGGCGCTTGGCAATACACCGCAAGGACGCGCGCTGCGCGGACCATCCCGGAC 426
QY 1174 GCCAGCAGCCCTCTATTTCGCCCAACACCGCATATGACCGATGCGGATATGGCATAAAG 1233
Db 425 CCGTTCCGCGGGCC---AGGCGCTCCCGACGATGCTGGCCACTGACCTCTCGCTGGG 369
QY 1234 GTAATTCGACCTATCGCGCTATCTCGCAAAAATTCATGGCCGATCTGAGTACTTCAAG 1293
Db 368 GTGGATCGGATCTATGAGCGGATCACGCTGCTGGTGGGAACACCCCGAGGAATGGCC 309
QY 1294 AAAACTTTCCGGAAGGGTGTTCAAGCTGACGACCGCTGACCTGGGCCCCGAAATCAAGT 1353
Db 308 GACGAGTTCGCGCAAGGCTGTACAAAGTGTACACCGAGACATGGGTCCCGTTGCGAGA 249
QY 1354 TACATCGCGCGGAGTGC CGGCAAGACCTGATTTGGCAAGACCCGATTTCCGCGAGGT 1413
Db 248 TACCTTGGCGCGCTGCTCCCAACGACAGACCTGCTGTGGCAGGATCCGCTCGCTGGGTC 189
QY 1414 AACACCGACTACTGCGAAGAAAGTGG-----TCAAGCAGAAAATTCGACAAAAGT 1461
Db 188 AGCCACGACCTCGTCCGCGAAGCGGAGATTCGACGCTTAAAGACCAAGTCTGCGATCG 129
QY 1462 GGCCTGAGCATTAGTGAGATGGTCTCCACCGCTTGGGACAGTGGCCGTACTTATCGCGGT 1521
Db 128 GATTGACTGCTACAGCTAGTTTCGACCGCATGGGCGGGCGGTCTGCTTCCGTGGT 69
QY 1522 TCCGATATGCGCGGGTGTAAACGTTGCCGATTCGCTTGGCCCCACAGAACAGATGG 1581
Db 68 AGCGACAAGCGCGCGGCGCAACGGTGTGCGCATCCGCTCGCAGCCCAAGTCGGGTGG 9
QY 1582 CAGGCCAA 1589
Db 8 GAGGTCAA 1

RESULT 8
US-09-941-193A-144
; Sequence 144, Application US/09941193A
; Publication No. US2003010873A1
; GENERAL INFORMATION:
; APPLICANT: BROW, MARY ANN D.
; LYAMICHEV, VICTOR I.
; OLIVE, DAVID M.
; TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
; PATHOGENS
; NUMBER OF SEQUENCES: 165
; CORRESPONDENCE ADDRESS:

Db 253 GTGGATCGATCTATGACGCGATCACGCGTCTGCTGGCTGGACACCCCGAGGAATTGGCC 312
QY 1294 AAAAATTCGGAAGCGGTGTTCAAGCTGACGACCGTACCTGGGCGCCGAAATCAAGT 1353
Db 313 GACGAGTTCCGCAAGCGCTGTTACAGCTGATCCACCGAGACATGGTCCCGTTCGGAGA 372
QY 1354 TACATCGCCCGGAGTCCGCGCAGACAGACCTGATTGGCAGACCCGATTCGCGCAGGT 1413
Db 373 TACCTTGGCGCGTGGTCCCAAGCAGACAGCTGCTGTGGCAGGATCCCGTCCCTGCGGTC 432
QY 1414 AACACCGACTACTCGGAAGAAGTGG-----TCAAGCAGAAAAATTGCACAAAGT 1461
Db 433 AGCCAGGACCTCGTCGGGAGCGGAGATGGCCAGCCTTAAGAGCCAGATCTTGCAATCG 492
QY 1462 GGCCTGAGCATAGTAGAGTGTCTCCACCGCTTGGGACAGTGGCCGCTACTATTCGCGGT 1521
Db 493 GGATTGACTGTCTCACAGCTAGTTTCGACCGCATGGGCGGCGCTCGTTCGTTGGTGT 552
QY 1522 TCCGATATCGCGCGGTGCTAAAGTGTCCCGCATTCGCTTGGCCCGCCACAGACAGTGG 1581
Db 553 AGCGACAAGCGCGCGCGCCCAACGGTGTGCGATCCCGCTCGACGCCACAAGTCCGGTGG 612
QY 1582 CAGGGCAA 1589
Db 613 GAGGTCAA 620

RESULT 10
US-09-941-193A-148/c
; Sequence 148, Application US/09941193A
; Publication No. US20030108873A1
; GENERAL INFORMATION:
; APPLICANT: BROW, MARY ANN D.
; OLIVE, DAVID M.
; TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
; PATHOGENS
; NUMBER OF SEQUENCES: 165
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 28-Aug-2001
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: FORS-01756
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 148:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 620 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 148:
US-09-941-193A-148

Query Match

6.1%; Score 137.2; DB 11; Length 620;

Best Local Similarity 54.3%; Pred. No. 4.7e-32;
Matches 330; Conservative 0; Mismatches 263; Indels 15; Gaps 2;
QY 994 GCAAGGCAAGCAACGCGTACCTTCGGGTATTCGAAGTGTCTGGACCAACCAACCCACG 1053
Db 605 GGAACCGTAAGGACGCGATCACACCGCATCGAGTCTGATGACGACACCCGAGG 546
QY 1054 AAATTCGATATGGGTATTTTCGACCTGCTGTTCGGCTACAAATGGAACACTGAAAAGAGT 1113
Db 545 AAATGACCAACAGTTTCTCGAGATCTCTGACGGTACGAGTCTGAGTCTGACGAGACG 486
QY 1114 CTTGCGGTGCGCCCAATTTGGAGCATTCGATCAAAAAGAAACAAAGCCGTTGAC 1173
Db 485 CTGCTGCGCTTGGCAATACACCCCAAGGACGGCGCGTCCGCGCACCATCCCGGAC 426
QY 1174 GCCAGCGACCCCTCTATTTCGCCACAACCCGATCATACCGCATCGCGATATGCGGATAAG 1233
Db 425 CGTTTCGGGGGCC---AGGGCGCTCCCGACGATGCTGGCCACTGACCTCTCGCTCGG 369
QY 1234 GTAAATCCGACCTATCGCGCTATCTGCGGAAAAATTCATGCGCGATCTCTGAGTACTTCAAG 1293
Db 368 GTGGATCCGATCTATGAGCGGATCACGGTCTGCTGGCTGGAACACCCCGAGGAATGGCC 309
QY 1294 AAACTTTTCGGAAGCGTGTTCAGCTGACGACCGCTGACCTGGCCCGCAATCAAGT 1353
Db 308 GACGAGTTCCGCAAGCGCTGTACAGCTGTATCCACCGAGACATGGTCCCTTTCGCGA 249
QY 1354 TACATCGCGCCGGAAGTCCCGCAGAACACTGTATTGGCAAGACCCGATTCGCGCAGGT 1413
Db 248 TACCTTGGCGCGCTGTGTCACAGCAGACCTGCTGTGCGAGATTCGCTCCCTCGGTC 189
QY 1414 AACACCGACTACTGCGAAGAAGTGG-----TCAAGCAGAAAAATTGCACAAAGT 1461
Db 188 AGCCACGACCTCGTCGGGGAAGCCGAGATTGCCAGCTTAAAGAGCCAGATCCGGGCATCG 129
QY 1462 GGCCTGAGCATAGTAGATGCTCTCCACCGCTTGGGACAGTGGCCGCTACTTATCGCGT 1521
Db 128 GGATTGACTGTCTCACAGCTAGTTTCGACCGCATGGGCGCGCTCGTTCGTTCCGTGCT 69
QY 1522 TCCGATATCGCGCGGTGCTAAAGTGTCCCGCATTCGCTTGGCCCGCCACAGACAGTGG 1581
Db 68 AGCGACAAGCGCGCGCGCCCAACGGTGTGCGATCCCGCTCGAGCCACAAAGTCCGGTGG 9
QY 1582 CAGGGCAA 1589
Db 8 GAGGTCAA 1

RESULT 11
US-09-941-193A-150/c
; Sequence 150, Application US/09941193A
; Publication No. US20030108873A1
; GENERAL INFORMATION:
; APPLICANT: BROW, MARY ANN D.
; LYAMICHEV, VICTOR I.
; OLIVE, DAVID M.
; TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
; PATHOGENS
; NUMBER OF SEQUENCES: 165
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 28-Aug-2001
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: FORS-01756
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 148:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 620 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 148:
US-09-941-193A-148

Db	433	AGCCAGACACTCGTGGCGGAAGCCGAGATTGGCAGGCTTAAAGACCCAGATCCGGGCGATCG	492
Qy	1462	GGCTGAGCATTAGTACAGATGGTCTCACCGCTTGGACAGATGCCCGTACTTATCGCGT	1521
Db	493	GGATTGACTCTCTCACAGTAGTTTCAGCGCATGGCGCGCGTCTGTCGTTCCTCGTGGT	552
Qy	1522	TCCGATATGCGCGCGGTGTAAAGTGGCCCGCATTCGCTTGGCCCCACAGAACAGTAGTG	1581
Db	553	AGCGACAAGCGCGCGCGCAACCGTGGTTCGATCCGCTGCAGCCACAAAGTCGCGGTGG	612
Qy	1582	CAGGCGAA 1589	
Db	613	GAGGTCAA 620	
RESULT 13			
US-09-940-925A-145			
; Sequence 145, Application US/09940925A			
; Publication No. US2003005438A1			
; GENERAL INFORMATION:			
; APPLICANT: BROW, MARY ANN D.			
; LYAMICHEV, VICTOR I.			
; OLIVE, DAVID M.			
; TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF			
; PATHOGENS			
; NUMBER OF SEQUENCES: 165			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: MEDLEN & CARROLL			
; STREET: 220 MONTGOMERY STREET, SUITE 2200			
; CITY: SAN FRANCISCO			
; STATE: CALIFORNIA			
; COUNTRY: UNITED STATES OF AMERICA			
; ZIP: 94104			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: PatentIn Release #1.0, Version #1.30			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/09/940,925A			
; FILING DATE: 10-Jun-2002			
; CLASSIFICATION: <unknown>			
; ATTORNEY/AGENT INFORMATION:			
; NAME: CARROLL, PETER G.			
; REGISTRATION NUMBER: 32,837			
; REFERENCE/DOCKET NUMBER: FORS-01756			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: (415) 705-8410			
; TELEFAX: (415) 397-8338			
; INFORMATION FOR SEQ ID NO: 145:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 620 base pairs			
; TYPE: nucleic acid			
; STRANDEDNESS: double			
; TOPOLOGY: linear			
; MOLECULE TYPE: DNA (genomic)			
; SEQUENCE DESCRIPTION: SEQ ID NO: 145:			
US-09-940-925A-145			
Query Match 6.0%; Score 135.6; DB 11; Length 620;			
Best Local Similarity 54.1%; Pred. No. 1.5e-31;			
Matches 329; Conservative 0; Mismatches 264; Indels 15; Gaps 2;			
Qy	994	GGCAAGCGCAAGCAACGCCGTGACCTCGGCTATCGAAGTCTCTGGACCACCAACCCACG	1053
Db	16	GGAACCGGTGAAGACCGGATCCACCGCGCATGAGGTCTGATGGAGCAACACCCGACG	75
Qy	1054	AAATTGATATGGCTATTTTCGACCTGCTTCGGCTACAAATGGGAACGTGAAAAAGAGT	1113
Db	76	AAATGGACAACAGTTTCCTCGAGATCCCTGACGGTACGAGTGGGAGCTGACGAAGAGC	135
Qy	1114	CTGCGCGGTGCCCACTTGGGAACCGATTGACATCAAAAAGGAAAAACGCCGTTGAC	1173

```

; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 147:
US-09-940-925A-147

Query Match
Best Local Similarity 6.08; Score 135.6; DB 11; Length 620;
Matches 329; Conservative 0; Mismatches 264; Indels 15; Gaps 2;

QY 994 GGAAGCAAGCAAGCAGCGGTGACCTCGGTATCAAGGTCTTGGACCAACACCCACG 1053
Db 605 GGAACCGGTAGGACGCGATCACCGGCGATCGAGTCTGATGAGCAACACCCGACG 546
QY 1054 AAATTCGATATGGCTATTTCCGACCTGCTGTTCGGCTACAAATTTGGAACTGAAAAGAGT 1113
Db 545 AAATGGCAACAGCTTTCTCGAGATCTGTACGGCTACGAGTGGAGCTGACCAAGAC 486
QY 1114 CTTGCCGCTGCCACCATTTGGAAACCGATGACATGACCGATGCGGATATGGCGATTAAG 1173
Db 485 CTTGCTGGCGTGGCAATACACCGCCAGGACGCGCGCGGTGCGCGACCATCCCGGAC 426
QY 1174 GCCAGCGACCCCTCTATTTCGCCCAACCCGATCATGACCGATGCGGATATGGCGATTAAG 1233
Db 425 CCGTTCGGCGGGC---AGGCGCTCCCGACGATGCTGCCACCTGACCTCTCGCGCGG 369
QY 1234 GTAATCCGACCTATCGCGTATCTGCGCAAAAATTTCAATTCATGCGCGATCTGAGTCTCAAG 1293
Db 368 GTGATCCGATCTATGAGCGGATCACCGCTGCTGGTGGCAACACCCGAGGATTTGCC 309
QY 1294 AAAACTTTCCGGAAGAGTGGCGGAGAGAGTGGCGGAGAGTGGCGGAGAGTGGCGGAGT 1461
Db 188 AGCCACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129
QY 1462 GGCGTACGATTTAGTGGATGGTCTCCACCGCTTGGACAGTGGCGCGCGCGCGCGCGCG 1353
Db 308 GACGATTTGCCAAGGCTGTGATCAGCTGATCCACCGAGACATGGGTCCGTTGCGGAGA 249
QY 1354 TACATCGCGCGGAGAGTGGCGGAGAGAGTCTGATTTGGCAAGACCGGATTTCCGCGAGT 1413
Db 248 TACCTTGGCGCGTGTGTCGCCCAAGCAGACCTGCTGTGGCAGGATCGGTCCTCGCGGTC 189
QY 1414 AACACCGACTACTGCGGAAGAGTGGG-----TCAGCAGAGAAAATTTGCACAAAGT 1461
Db 188 AGCCACAGCGCTGTCGGCGAAGCGAGATTTGCCAGCCTTAAGAGCCAGATCTCGGCATCG 129
QY 1462 GGCGTACGATTTAGTGGATGGTCTCCACCGCTTGGACAGTGGCGCGCGCGCGCGCG 1353
Db 128 GGATTTGACTCTCAGACAGTATTTCCACCGCATTTGGCGCGCGCGCGCGCGCGCGGT 69
QY 1522 TCCGATATGCGCGCGGTGCTAAAGGTGCGCGGATTTGCTTTGGCGCGCGCGCGCGCGGT 1581
Db 68 AGCAGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTG 9
QY 1582 CAGGGCAA 1589
Db 8 GAGGTCAA 1

```

```

RESULT 15
US-09-940-925A-149/c
; Sequence 149, Application US/09940925A
; Publication No. US20030054338A1
; GENERAL INFORMATION:
; APPLICANT: BROW, MARY ANN D.
; LYAMICHEV, VICTOR I.
; OLIVE, DAVID M.
; TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
; PATHOGENS
; NUMBER OF SEQUENCES: 165
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/940,925A
; FILING DATE: 10-Jun-2002
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: FORS-01756
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 149:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 620 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 149:
US-09-940-925A-149

```

```

Query Match
Best Local Similarity 6.08; Score 135.6; DB 11; Length 620;
Matches 329; Conservative 0; Mismatches 264; Indels 15; Gaps 2;

QY 994 GGAAGCAAGCAAGCAGCGGTGACCTCGGTATCAAGGTCTTGGACCAACACCCACG 1053
Db 605 GGAACCGGTAGGACGCGATCACCGGCGATCGAGTCTGATGAGCAACACCCGACG 546
QY 1054 AAATTCGATATGGCTATTTCCGACCTGCTGTTCGGCTACAAATTTGGAACTGAAAAGAGT 1113
Db 545 AAATGGCAACAGCTTTCTCGAGATCTGTACGGCTACGAGTGGAGCTGACCAAGAC 486
QY 1114 CTTGCCGCTGCCACCATTTGGAAACCGATGACATGACCGATGCGGATATGGCGATTAAG 1173
Db 485 CTTGCTGGCGTGGCAATACACCGCCAGGACGCGCGCGGTGCGCGACCATCCCGGAC 426
QY 1174 GCCAGCGACCCCTCTATTTCGCCCAACCCGATCATGACCGATGCGGATATGGCGATTAAG 1233
Db 425 CCGTTCGGCGGGC---AGGCGCTCCCGACGATGCTGCCACCTGACCTCTCGCGCGG 369
QY 1234 GTAATCCGACCTATCGCGTATCTGCGCAAAAATTTCAATTCATGCGCGATCTGAGTCTCAAG 1293
Db 368 GTGATCCGATCTATGAGCGGATCACCGCTGCTGGTGGCAACACCCGAGGATTTGCC 309
QY 1294 AAAACTTTCCGGAAGAGTGGCGGAGAGAGTGGCGGAGAGTGGCGGAGAGTGGCGGAGT 1461
Db 188 AGCCACAGCGCTGTCGGCGAAGCGAGATTTGCCAGCCTTAAGAGCCAGATCTCGGCATCG 129
QY 1462 GGCGTACGATTTAGTGGATGGTCTCCACCGCTTGGACAGTGGCGCGCGCGCGCGCG 1353
Db 308 GACGATTTGCCAAGGCTGTGATCAGCTGATCCACCGAGACATGGGTCCGTTGCGGAGA 249
QY 1354 TACATCGCGCGGAGAGTGGCGGAGAGAGTCTGATTTGGCAAGACCGGATTTCCGCGAGT 1413
Db 248 TACCTTGGCGCGTGTGTCGCCCAAGCAGACCTGCTGTGGCAGGATCGGTCCTCGCGGTC 189
QY 1414 AACACCGACTACTGCGGAAGAGTGGG-----TCAGCAGAGAAAATTTGCACAAAGT 1461
Db 188 AGCCACAGCGCTGTCGGCGAAGCGAGATTTGCCAGCCTTAAGAGCCAGATCTCGGCATCG 129
QY 1462 GGCGTACGATTTAGTGGATGGTCTCCACCGCTTGGACAGTGGCGCGCGCGCGCGCG 1353
Db 128 GGATTTGACTCTCAGACAGTATTTCCACCGCATTTGGCGCGCGCGCGCGCGGT 69
QY 1522 TCCGATATGCGCGCGGTGCTAAAGGTGCGCGGATTTGCTTTGGCGCGCGCGCGCGGT 1581
Db 68 AGCAGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTG 9
QY 1582 CAGGGCAA 1589
Db 8 GAGGTCAA 1

```

Search completed: October 8, 2003, 19:55:23
Job time : 586.569 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2003, 07:49:13 ; Search time 4588.37 Seconds
(without alignments)
11726.187 Million cell updates/sec

Title: US-09-884-889-5
Perfect score: 2262
Sequence: 1 atgataacgcacccgtga.....gttcgacgtcgctgtaa 2262

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrt:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	285.2	12.6	1102	28	AF075865
2	261.4	11.6	707	13	BQ751801
3	259.8	11.5	762	13	BQ752115
4	257.8	11.4	562	10	BG370112
					Lma123 V8

5	246.2	10.9	742	13	BQ751881
6	236	10.4	639	13	BQ751512
7	234.8	10.4	606	13	BQ751387
8	233.2	10.3	729	9	AW179968
9	224.6	9.9	737	29	CNS01HHH
10	221.6	9.8	663	12	BM865076
11	217.6	9.6	633	13	BQ752114
12	217	9.6	757	14	CD425096
13	208.4	9.2	573	9	AW180683
14	206.6	9.1	701	14	CD422961
15	201.8	8.9	712	13	BU062489
16	201.2	8.9	656	13	BU062488
17	201	8.9	741	14	CD428471
18	197.6	8.7	660	13	BQ751388
19	197	8.5	666	28	AQ400148
20	193.4	8.5	666	28	AQ324744
21	192.2	8.5	918	28	AF075904
22	185.8	8.2	700	9	AW180906
23	185.4	8.2	646	28	AQ399314
24	184.8	8.2	571	9	AW180185
25	183	8.1	598	14	CD426092
26	183	8.1	710	13	BU065388
27	181.6	8.0	613	13	BQ751511
28	181.4	8.0	600	9	AW180270
29	179.2	7.9	673	9	AW180945
30	177.2	7.8	531	13	BQ143330
31	175.6	7.8	465	10	BG278355
32	169.8	7.5	945	14	CD459390
33	167.6	7.4	546	13	BQ142683
34	164.6	7.3	786	28	BH403595
35	163.6	7.2	588	14	CD457487
36	163.6	7.2	873	14	CD377714
37	159.6	7.1	745	14	CD425011
38	158	7.0	757	14	CD427691
39	156	6.9	621	12	BM612625
40	151.6	6.7	481	13	BQ751880
41	149.2	6.6	514	14	CD428369
42	149.2	6.6	718	12	B1750328
43	147	6.5	371	12	BM865392
44	142.6	6.3	358	28	A2578961
45	137.6	6.1	565	28	BH401331

ALIGNMENTS

RESULT 1

AF075865

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

AF075865 1102 bp DNA linear GSS 29-AUG-2000
AF075865 Salmonella typhimurium LT2, Lambda DASH II Salmonella
typhimurium genomic clone 1079-T7, genomic survey sequence.

AF075865
AF075865
GSS.
Salmonella typhimurium
Salmonella typhimurium
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.

1 (bases 1 to 1102)
Wong,R.M.-Y., Wong,K.K., Benson,N.R. and McClelland,M.
Sample sequencing of a Salmonella typhimurium LT2 lambda library:
Comparison to the Escherichia coli K12 genome
FEMS Microbiol. Lett. 173 (2), 411-423 (1999)

99243757
10227170
Contact: McClelland M
Molecular Biology
Sidney Kimmel Cancer Center
3099 Science Park Road, San Diego, CA 92121, USA
Email: mclelland@lifsci.sdsu.edu
Class: shotgun.
Location/Qualifiers
1. .1102

FEATURES

source

db

Db
160 CCGCATCAAGAACCGCGACTGGTGGCCGGATGCGGTGAAGCTCAACATCCTCCGCCAGCA 219

TITLE Expressed sequence tags of *Leptosphaeria maculans*, cause of
blackleg disease of oilseed *Brassica*
JOURNAL Unpublished
COMMENT Contact: Alexander Idnurm
Barbara J. Howlett (howlett@unimelb.edu.au)
School of Botany
The University of Melbourne, Parkville, VIC, 3010, Australia
Tel: 61 3 8344 5056
Fax: 61 3 9347 5460
Email: a.idnurm@botany.unimelb.edu.au
Insert length: 2700 Std Error: 0.00
Seq primer: T3 Universal Primer.
Location/Qualifiers
1. .562
/organism="Leptosphaeria maculans"
/mol_type="mRNA"
/strain="M1"
/db_xref="taxon:5022"
/clone_lib="V8 mycellial library"
/note="Library prepared from mycellium grown in 10%
Campbell's V8 juice."
BASE COUNT 116 a 178 c 157 g 111 t
ORIGIN
Query Match 11.4% Score 257.8; DB 10; Length 562;
Best Local Similarity 66.4%; Pred. No. 2.1e-59;
Matches 370; Conservative 0; Mismatches 187; Indels 0; Gaps 0;
QY 112 GGTAAATGCTGTCATGCACGGTGGTAAACCTCGACCGGTACTTCACAAAGATGG 171
Db 6 GGGAGAGTCTCCCGTAAAGACGGCAATGTGCGCGCGCGGTACTAGGAATGTGCACTGG 65
QY 172 TGGCGGAAGGGTTGAACCTGGATATTTGCATCAGCAAGATCGCAATCAGACCCGATG 231
Db 66 TGCCCAACAGCTGAGGACCAACATCCTCGCCAGCAGCAGTATCCCGCCAAACCCCTCTC 125
QY 232 GATCCGGATTCAACTACCTGGAAGAAGTAGCAAGCTCGATTTTCGACCGCTCAAGAAA 291
Db 126 GGCCCCGACTTGACATATGCGGAGGAGCTTCAAGAAGCTCGACTACGAGGCTTGAAGAG 185
QY 292 GATGTCCACCGTTGATGACCGATAGCCAGAGTGGTGGCCGCTGACTGGGGCGACTAC 351
Db 186 GATCTCAATCCCTTGATGACCGACTCCAGGACTGGTGGCGTGGCGACTTTGGCCACTAC 245
QY 352 GCGCGTTTGATGATCCGTATGGCTTGGCTCGCTGCGCACTACCGCTATTGCTGATGGC 411
Db 246 GGTGGTTCTTCATCCGATTGGCATGCGACGGCCCGCTACCTATCGTACAGGATGGC 305
QY 412 CGTGGGGCGCGGTGATACCGGAAGCCAGCGCTTTGACCCGCTCAACTCCTGGCCGACAAC 471
Db 306 CGTGGTGGAGGTGGCGAAGGACAACAGCGCTTTGCCCCCTCTCAACTCATGGCCGACAAT 365
QY 472 GTACGCTGATAAAGGCCCGCTGCTGCTGGCCGATCAAGAAGAAGTACGGGCAACAAA 531
Db 366 GTCTCGCTTGACAAGGCTCGCTCGCCTTTGTGGCCCATCAAGCAAAAAGTATGGCAACA 425
QY 532 ATCAGCTGGCAGACCTGATGATCTGGCTGGCACCCTGGCTATGATCCATGGGCTTA 591
Db 426 ATCTCTATGCCGACCTCATGCTCTGGCTGGCAACGCTCGCCCTCGAGTCCATGGGTTC 485
QY 592 CTTGCTTACGGCTTCTTTTGGCGCGCGGTGATATTTGGGAACCCGAAAAAGATATCTAC 651
Db 486 CCCACATTCGGTTTCGCCGCTGGCCGTACGACACCTTCACAGTCTGACGAGTCGGTCTAC 545
QY 652 TGGGGTGACAAAAGA 668
Db 546 TGGGGAGCGGAGACAGA 562

DEFINITION EST632444 DSCT Colletotrichum trifolii cDNA clone pDSCT8-27, mRNA sequence.

ACCESSION BO751881

VERSION BO751881.1 GI:21907286

KEYWORDS EST.

SOURCE Colletotrichum trifolii
Colletotrichum trifolii
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetes Incertae sedis; Phyllachorales; Phyllachoraceae;
mitosporic Phyllachoraceae; Colletotrichum.
1 (bases 1 to 742)

REFERENCE Samac,D.A., Dickman,M., Town,C.D., Van Aken,S., Utterback,T.,
AUTHORS Cheung,F. and Fraser,C.M.
TITLE ESTs from mycelia of Colletotrichum trifolii race 1
JOURNAL Unpublished
COMMENT Other_ESTs: EST632443
Contact: Deborah A. Samac
Department of Plant Pathology
University of Minnesota
495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA
Tel: 612 625 1243
Fax: 651 649 5058
Email: debys@umn.edu
TIGR sequence name: MTSAH27IV More information is available at:
www.medicago.org
Seq primer: (gta ata Cga CtC Act ACA ggg C).

FEATURES
source
 1..742
 /organism="Colletotrichum trifolii"
 /mol_type="mRNA"
 /strain="race 1"
 /db_xref="taxon:5466"
 /clone="pDSCT8-27"
 /tissue_type="mycelia"
 /dev_stage="Young, actively growing mycelia (3 days after inoculation) grown in liquid culture (cutin minimal medium containing 2glucose)."
 /lab_host="DH5alpha"
 /clone_lib="DSCN"
 /note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2: EcoRI; Isolate: 28p2 ; cDNA was prepared from polyA+ enriched RNA The cDNA was ligated into Lambda gt10 from Stratagene and packaged using Gigapack packaging extracts. An aliquot of the amplified library was used to transduce E. coli Y1090 and phage DNA was purified from a liquid lysate. The cDNA inserts were gel purified after EcoRI digestion and ligated into pBluescript SK+. Aliquots of the ligation were used to transform E. coli DH5alpha which were plated onto medium with X-gal for selection of recombinants."

	BASE COUNT	151 a	259 c	212 g	120 t
ORIGIN					
Query Match		10.9%;	Score 246.2;	DB 13;	Length 742;
Best Local Similarity		67.9%;	Pred.No. 3.6e-56;		
Matches 359;	Conservative	0;	Mismatches 168;	Indels 2;	Gaps 1;
QY	150	CGGTACTTCACAAGAAGATTGGTGCCGGAAGGGTTGAACCTGGATATTTTGCATCAGCA	209		
Db	160	CGGCATCAAGAACCAGCGGACTGGTGGCGGATGGCGTGAAGCTCAACATCCTCGGCCAGCA	219		
QY	210	AGATCCAAATCAGACCCCATGGATCGGATTCAGTTCACCTACCGTGAAGAAGTAGCCGACGCT	269		
Db	220	CAGCCCCGTACCAACCCGCTGGCCGAGGACTTTGACTACGTGCGCGCCTTCAAGTCGCT	279		
QY	270	CGATTTTCGACCGCCTCAAGAAGATGTCACCGGGTTGATGACCGATAGCCAGAGTGGTG	329		
Db	280	CGACTACGAGGGCGTCAAGAAGAACCTCACGGCCCTGATGACCGACTCCAGGACTGGTG	339		
QY	330	GCCCCTGACTGGGGGCATACGCGGGTTTGATGATCCGATGGCTTGGCATCCTCCGCTGG	389		
Db	340	GCCTGCCGACTTTGGCCACTACGCGGGTGTGTTTCATCCGATGGCCCTGTCACAGCSCCG	399		

```

QY 390 CACCTACCGTATTGCTGATGCCGTG--GGGCGGTGTGTACCGGAAGCGACGGCTTTGCA 447
Db 400 CACGTACCGGTTTACGACGGAGCGGAGGGGGTGGGAGAGGCGCAGCAACGCTTCGCA 459
QY 448 CGGCTCAACTCCTGCGCGGACACGCTCAGCCCTGGGATAAAGCGCGCTGTGTTGGCGG 507
Db 460 CGGCTCAACAGTGGCGCGGACNATGTACCTTCGACAGGCGCGCTGTGTTGGCGG 519
QY 508 ATCAAGAAGATACGGCAACAAATACAGTGGGAGACCTGTATGTTTGGCTGGCAC 567
Db 520 ATCAAGCAAAAGTACGGCAACAAGATCTCGTGGCGGACCTCATGATCCTGGCGGCAAC 579
QY 568 GTGGCTTATGACTCATGGCTTACTCTGCTTACGGCTTCTCTTTCGGCGGCGTGCATATT 627
Db 580 GTGGCGCTCGAGTCCATGGGTTCCAGAGCGCGCGCTTCTCGGAGGCGCGTCCCGACAC 639
QY 628 TGGGAACCGGAAAAATATCTACTGGGTGACGAAAAAGAGTGGCTGG 676
Db 640 TGGGAACCGGAGGCTCGTCTACTGGGCGGCGGAGACACCTGGCTGG 688

RESULT 6
BQ751512
LOCUS BQ751512
DEFINITION EST632075 DSCT Colletotrichum trifolii cDNA clone pDSCT5-66, mRNA
sequence.
ACCESSION BQ751512
VERSION BQ751512.1 GI:21906917
KEYWORDS Colletotrichum trifolii
SOURCE Colletotrichum trifolii
ORGANISM Colletotrichum trifolii
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetes incertae sedis; Phyllachorales; Phyllachoraceae;
mitosporic Phyllachoraceae; Colletotrichum.
1 (bases 1 to 639)
Samad,D.A., Dickman,M., Town,C.D., Van Aken,S., Utterback,T.,
Cheung,F. and Fraser,C.M.
ESTs from mycelia of Colletotrichum trifolii race 1
Unpublished
Other_ESTs: EST632074
Contact: Deborah A. Samad
Department of Plant Pathology
University of Minnesota
495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA
Tel: 612 625 1243
Fax: 651 649 5058
Email: debbyspuccini.crl.umn.edu
TIGR sequence name: MTSAE66TV More information is available at:
www.medicago.org
Seq primer: (9tA AtA CgA CUC Act AtA 99g C).
Location/Qualifiers
1. .639
/organism="Colletotrichum trifolii"
/mol_type="mRNA"
/strain="race 1"
/db_xref="taxon:5466"
/clone="pDSCT5-66"
/tissue_type="mycelia"
/dev_stage="young, actively growing mycelia (3 days after
inoculation) grown in liquid culture (cutin minimal medium
containing 2%glucose)."
/lab_host="DH5alpha"
/clone_lib="DSCT"
/notes="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:
EcoRI; Isolate: 2sp2 ; cDNA was prepared from polyA+
enriched RNA The cDNA was ligated into Lambda gtl1 from
Stratagene and packaged using Gigapack packaging extracts.
An aliquot of the amplified library was used to transduce
E. coli Y1090 and phage DNA was purified from a liquid
lysate. The cDNA inserts were gel purified after EcoRI
digestion and ligated into pBluescript SK+. Aliquots of
the ligation were used to transform E. coli DH5alpha which
were plated onto medium with X-gal for selection of

```

```

BASE COUNT 131 a 222 c 176 g 110 t
ORIGIN
Query Match 10.4%; Score 236; DB 13; Length 639;
Best Local Similarity 68.5%; Pred. No. 2.1e-53;
Matches 326; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
QY 150 CGGTACTTCCAAAGAGATTGTTGGTCCGGAAGGGTTGAACCTTGGATATTTTCATCAGCA 209
Db 164 CGGCATCAAGAACCGGACTGTTGGCGGAGTCCGCTGAAGCTCAACATCTCCGCCAGCA 223
QY 210 AGATCGCAATCAGACCCGATGGATCCGGATTCAACTACCTGGAAGAAGTACGCAAGCT 269
Db 224 CACGCCCGTCAACAAACCGCTGGGCAGGACTTTGACTACGTCGCGCGCTTCAAGTCGCT 283
QY 270 CGATTTCGACGGCTGAAGAAGATGTCACCGTTGATGACCGATAGACCAAGAGTGGTG 329
Db 284 CGACTACGAGGCGTCAAGAAGACCTCAGGCCCTGATGACCGACTCCCGAGACTGGTG 343
QY 330 GCCCGTGACTGGGGCACCTACGGCGGTTTGTATGATCCGTATGGCTTGGCACTCCGCTGG 389
Db 344 GCCTCGGACTTTGGCCACTACGGCGGTCTGTTCATCCGCATGGCTGGCAGACGCCGG 403
QY 390 CACCTACCGTATTGCTGATGGCGGTGGGGCGGTGTACCGGAAGCGCTTTGGCAC 449
Db 404 CACGTACCGAGTTCACGACGGAGCGGAGTGGTGAGAGGCCCAACGCTTCGCACC 463
QY 450 GCTCAACTCTGCGCGGACACGTCAGCTGATGAAGCGCGCTCTGCTGCGCGAT 509
Db 464 GCTCAACAGTGGCGGCAATGTACGCTCAGCAAGGCCCTCGGCTCTGTGGCCCAT 523
QY 510 CAAGAAGAGTACGGCAACAAATCAGCTGGCGACACCTGATGATCTGGCTGGCACCGT 569
Db 524 CAAGCAAAAGTACGGCAACAAGATCTCGTGGCGGACCTCATGATCTGGCGGCAACGT 583
QY 570 GCCTATGAGTCCATGGGCTTACCTGCTTACGGCTTCTCTTTCGGCGGCTGATA 625
Db 584 GGCCTCGAGTCCATGGGTTTCCAGACGCGGCGCTTCTCCGAGGCGGCTCCCGACA 639

RESULT 7
BQ751387
LOCUS BQ751387
DEFINITION EST631950 DSCT Colletotrichum trifolii cDNA clone pDSCT4-87, mRNA
sequence.
ACCESSION BQ751387
VERSION BQ751387.1 GI:21906792
KEYWORDS Colletotrichum trifolii
SOURCE Colletotrichum trifolii
ORGANISM Colletotrichum trifolii
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetes incertae sedis; Phyllachorales; Phyllachoraceae;
mitosporic Phyllachoraceae; Colletotrichum.
1 (bases 1 to 606)
Samad,D.A., Dickman,M., Town,C.D., Van Aken,S., Utterback,T.,
Cheung,F. and Fraser,C.M.
ESTs from mycelia of Colletotrichum trifolii race 1
Unpublished
Other_ESTs: EST631951
Contact: Deborah A. Samad
Department of Plant Pathology
University of Minnesota
495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA
Tel: 612 625 1243
Fax: 651 649 5058
Email: debbyspuccini.crl.umn.edu
TIGR sequence name: MTSAD87TK More information is available at:
www.medicago.org
Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
Location/Qualifiers
1. .606
/organism="Colletotrichum trifolii"

```

FEATURES
source

/mol_type="mRNA"
 /strain="race 1"
 /db_xref="taxon:5466"
 /clone="pDSC74-87"
 /tissue_type="mycelia"
 /dev_stage="Young, actively growing mycelia (3 days after inoculation) grown in liquid culture (cutin minimal medium containing 2%glucose)."
 /lab_host="DH5alpha"
 /clone_lib="DSC7"
 /note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2: EcoRI; isolate: 2sp2 ; cDNA was prepared from polyA+ enriched RNA The cDNA was ligated into lambda gtl from Stratagene and packaged using Gigapack packaging extracts. An aliquot of the amplified library was used to transduce E. coli Y1090 and phage DNA was purified from a liquid lysate. The cDNA inserts were gel purified after EcoRI digestion and ligated into pBluescript SK+. Aliquots of the ligation were used to transform E. coli DH5alpha which were plated onto medium with X-gal for selection of recombinants."
 BASE COUNT 124 a 214 c 172 g 96 t
 ORIGIN

Query Match 10.48; Score 234.8; DB 13; Length 606;
 Best Local Similarity 68.8%; Pred. No. 4.4e-53;
 Matches 337; Conservative 0; Mismatches 152; Indels 1; Gaps 1;
 QY 150 CGGTACTTCCAAACAGATTGGTGGCGGAAGGTTGAACCTGGATTTTGGCATCAGCA 209
 DB 117 CGGATCAAGAACCGGACGTGGTGGCGGATGGCTGAAGCTCAACATCTCCGCCAGCA 176
 QY 210 AGATCGCAATCAGACCGGATGATCGGATTTCAACTACCTGCGTGAAGAGTACGCAAGCT 269
 DB 177 CACGCCGTACCAACCCGCTGGCGAGGACTTGACTAGCTGCGCCCTTCAAGTCGT 236
 QY 270 CGATTTGCGAGCGGTGAAGAAAGATGCCAGCGTTGATGACCGATAGCCAGTGGTG 329
 DB 237 CGACTAGAGGGGCTCAAGAAGACCTCACGGCCCTGATGACGAGCTCCCGAGGACTGGT 296
 QY 330 GCCCGTGTACTGGGGCACTACGCGGTTTGTATGATCCGTATGCTTGGGCACATCCGCTGG 389
 DB 297 GCCTGCGGATTTGGCCACTACGGCGTCTGTCTATCCGATGCGCTGGCACAGCGCCGG 356
 QY 390 CACCTACCGTATTGCTGATGGCGTGG -GGGCGGTGGTACCGGAAGCAGCGCTTTGCAC 448
 DB 357 CAGGTACCGAGTTCACGACGCGGAGGTGGTGGAGAGGCGCAACACGCTTCGCAC 416
 QY 449 CGCTCACTCTGGCGGACACGTCAGCGTGGATGAAGCGCGCGCTCTGCTGTGGCCGA 508
 DB 417 CGCTCAACAGCTGGCGGACAAATGTACAGCTCGACAAAGGCCGCTGCTGTGGCCCA 476
 QY 509 TCAAGAAGAGTACGGCAACAAATCAGCTGGGCGACGCTGATGATTCGTGGCTGGGACCG 568
 DB 477 TCAAGCAAAAGTACGGCAACAGATCTGTGGCGGACCTCATGATCTGCGCGGCAACG 536
 QY 569 TGCGTTATGATGATCCATGGGTTACCTGCTTACGGCTTCTCTTTTCGGCGCGCTCGATATT 628
 DB 537 TGGCCCTCGAGTCCATGGGTTTCAGACGCGCGGCTTCTCCGGAGGCGGTCGCCGACCT 596
 QY 629 GGGAAACCGA 638
 DB 597 GGGAAACCGA 606

RESULT 8
 AW179968 729 bp mRNA linear EST 17-NOV-1999
 LOCUS Mga0020f Mga Library Mycosphaerella graminicola cDNA clone Mga0020
 DEFINITION 5' similar to catalase-peroxidase, mRNA sequence.
 ACCESSION AW179968
 VERSION AW179968.1 GI:6447163
 KEYWORDS EST.

SOURCE ORGANISM

Mycosphaerella graminicola
 Mycosphaerella graminicola
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes et
 Chaetothyriomycetes incertae sedis; Mycosphaerellaceae;
 Mycosphaerella.

REFERENCE

1 (bases 1 to 729)

AUTHORS

Koon,J.P.R., Bailey,A.M. and Hargreaves,J.A.

TITLE

A group of expressed cDNA sequences from the wheat fungal leaf
 blotch pathogen, Mycosphaerella graminicola (Septoria tritici)

JOURNAL

Fungal Genet. Biol. 29 (2), 118-133 (2000)

MEDLINE

20374020

PUBMED

10919380

COMMENT

Contact: Hargreaves JA

Cell Biology Department

IACR-Long Ashton Research Station

Long Ashton, Bristol, BS41 9AF, UK

Tel: +44(0)1275 392181

Fax: +44(0)1275 394281

Email: John.hargreaves@bbsrc.ac.uk

Insert Length: 2600 Std Error: 0.00

Seq primer: M13 reverse.

Location/Qualifiers

1..729

/organism="Mycosphaerella graminicola"

/mol_type="mRNA"

/strain="Strit"

/db_xref="taxon:54734"

/clone="Mga0020"

/clone_lib="Mga Library"

/note="Vector: pSPORI1; Library constructed from cultures
 utilizing ammonium ions as a source of nitrogen"

BASE COUNT 159 a 217 c 203 g 150 t

ORIGIN

Query Match 10.38; Score 233.2; DB 9; Length 729;

Best Local Similarity 65.2%; Pred. No. 1.3e-52;

Matches 343; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

QY 151 GGTACTTCCAAACAAAGATTGGTGGCGGAAGGTTGAACCTGGATATTTTGCATCAGCA 210

DB 92 GTTACCCGCAACAGGAGCTGGTGGCCAAATGAGCTGAACCAAGATCTTGGCCGACAC 151

QY 211 GATCGCAAAATCAGACCGCGATGGATCCGGATTTCAACTACCGTGAAGAAGTACGCAAGCTC 270

DB 152 ACCGCTGCTACCGACCCATTTCGGCAAGCAGTTCGACTACCCAGCGCATTCAGAGAGCTTG 211

QY 271 GATTTCCAGCGCTGAGAAAGATGTCCACGCGTTGATGACCGATACCCAAAGTGTGGTGG 330

DB 212 GATTACAATGGCTCAAGAAGGACTTGAACGACCTCATGACCGACTCGAAGGACTTCTGG 271

QY 331 CCGCGTGACTGGGGCACTACGCGGTTTGTATGATCGTATGGCTTGGCACTCGCGTGGC 390

DB 272 CGAGCAGACTTTGGCCACTACGGTGGTTCTTCTGTCGCGATGGATGGCACAGCGCTGGT 331

QY 391 ACCTACCGTATTGCTGATGCCCGTGGGGCGGTGGTACCGGAAGCCAGCCGCTTTGCACCG 450

DB 332 ACCTACCGTCTGCTGATGGTCTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 391

QY 451 CTCAACTCTTGGCGGACACAGCTCAGCTGGATGAAGCGCGCGTCTGCTGTGGCGGATC 510

DB 392 CTCAACTATGGCGGACAAACGCAATCTCGAAGGCTCGCGCGCTGTGTGGTGGCAATC 451

QY 511 AAGAAGAGTACGCAACAAATCAGCTGGCGAGACTGATGATCTTGGCTGCACCGCTG 570

DB 452 AAGCAAAAGTACGCAACAAAGATCTATGGCGGACCTGTCTCTCTCACTGGAATGTC 511

QY 571 GCTTATGAGTCCATGGGCTTACCTGCTTACCGGCTTCTCTTTCGGCGCGCTGCGATATTGG 630

DB 512 GCCATTGAGTCTATGGTCTGCGCACGTTTGGTTTTCGCGGTTGGCGGTGCGCGACATGG 571

QY 631 GAACCCGAAAAAGATATCTACTGGGTGACGAAAAAGAGTGGCTGG 676

DB 572 GAGCGGATGATCTCCGTATATCTGGGTGGTGGAGACCACTGTGTTGG 617


```

||||| | ||||| | || | ||||| | || | ||||| | || | ||||| | || | ||||| | || |
Db 202 TACGAGGCGAGGCGGAGGTCGGCGGAGAGAGTGGAGCGGCGGAGCGGTCGACCTC 143
QY 2125 GTATTGTGTCGAACCTGCTACTGCGCTCTTACGCGAGAGGTGTACGCCCGGAGCAGCAATAC 2184
Db 142 GTGTCGGGTCCGAGCGGAGCTGGTGGGATTTCCGAGAGGTGTATGCCAGCGGATGGC 83
QY 2185 GCGGAGAGTTCGTGAGAGACTTCGTCGCCGCTGGACCAAGTGTATGAACCGCGACCGT 2244
Db 82 GCGGAGAGTTCGTGAGGAGCTTTGTGAGCGCGTGGGATAAGGTGATGAACCTGGATCGG 23
QY 2245 TTCGACGT 2252
Db 22 TTCGACCT 15

CD425096 757 bp mRNA linear EST 02-JUN-2003
SAL10_G08.g1_A002 Salicylic acid-treated seedlings Sorghum bicolor
CDNA clone SAL10_G08_A002 5', mRNA sequence.
CD425096
ACCESSION CD425096.1 GI:31331359
VERSION EST.
KEYWORDS Sorghum bicolor (sorghum)
SOURCE Sorghum bicolor
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 757)
Cordonnier-Pratt,M.-M., Wentzel,V., Suzuki,Y., Sugano,S., Klein
,R.R., Liang,C., Sun,F., Sullivan,R., Shah,M., Salzman,R., Chua Tan
,N., Gonzalez,M., Lane,S., Miller,V., Nanda,P., Olaseinde,O.,
Eastman,A. and Pratt,L.H.
An EST database from Sorghum: salicylic acid-treated seedlings
Unpublished
Other_ESTs: SAL10_G08.bl_A002
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in
the Human Genome Center, University of Tokyo Institute of Medical
Science; plant material and RNA prepared at Texas A & M University;
sequencing done in the Laboratory for Genomics and Bioinformatics,
University of Georgia. Sequence ends have been trimmed to exclude
vector and regions below Phred quality 16. Three-prime sequences
are presented as their reverse complement and have been trimmed to
exclude polyA.
Seq primer: Sug5 (CTTCTGCTCTAAAGCTGGC).
Location/Qualifiers
1..757
/organism="Sorghum bicolor"
/mol_type="mRNA"
/cultivar="BTx623"
/db_xref="taxon:4558"
/clone="SAL10_G08_A002"
/lab_host="DH10B-T1 phage-resistant E. coli"
/clone_lib="Salicylic acid-treated seedlings"
/notes="Vector: pME18S-FL3; Site_1: XhoI; Site_2: XhoI; The
library was prepared from polyA+ RNA from seedlings grown
in hydroponic culture. At 8 days of age, medium was
supplemented with 1 mM salicylic acid (SA). Roots and
shoots were harvested after 27 and 72 hr and material from
both time points was combined prior to RNA isolation.
Double-stranded cDNA was cloned unidirectionally into
different DraIII sites of the pME18S-FL3 vector (5-prime
DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG
). XhoI excises the cDNA insert."
143 a 234 c 211 g 169 t

```

```

ORIGIN
Query Match 9.6%; Score 217; DB 14; Length 757;
Best Local Similarity 63.9%; Pred. No. 3.6e-48;
Matches 328; Conservative 0; Mismatches 185; Indels 0; Gaps 0;
QY 150 CCGTACTTCCAAACAAGATTTGGTGGCCGGAAGGGTTGAACCTGGATATTTTGCATCAGCA 209
Db 156 CCGTACTCGATCCAGGAGCTTTTGGCCCTTGTCTTTGAGACTGGATGTGTTCGAGCAGTT 215
QY 210 AGATCGCAAAATCAGACCGATGGATCCGATTTCAACTACCGTGAAGAAGTACGCAACGT 269
Db 216 TTCGCTCAGTATAACCCATTGGGTGCTGATTTTCACACACCGAGGCTTTCAAGTCTCT 275
QY 270 CGATTTCGACGCGTGAAGAAGATGTCACGCGTTGATACCGATAGCCAGAGTGGTG 329
Db 276 GGACTTTGCTGCATTGAAGAAGGACCTCAACGCGCTTCTCACTGATTCTCAGGATTGGTG 335
QY 330 GCCCGCTACTGGGGGCGACTACGGGGGTTTGNATGATCCGTATGGCTTGGCACTCCGCTGG 389
Db 336 GCCTGCTGACCATGTAACACTATGTTGGTCTCTTCAATTCGTATGTCATGTCACGACGCTGG 395
QY 390 CACCTACCGTATTGCTGATGGCCGTGGGGGGTGTACCGGAAGCCAGCGCTTTGCACC 449
Db 396 TACATACCGCGCAACGGATGGCCGAGTGGCGCCGGAATGGGTCAACACGATTGCTCC 455
QY 450 GGTCAACTCCTGGCCGGACACGCTCAGCTGATGAATAAGCGCCGCTCTGCTGGCCGAT 509
Db 456 TCTCGACAGCTGGCCCGACACCAACAGCTGGACAAGCCCGCTGCTGCTGCGCCCAT 515
QY 510 CAGAAGAAGTACGGCAACAATAATCAGCTGGCGACAGCTGATGTTCTGGCTGGCACCGT 569
Db 516 CAAGCAAAAGTACGGCAACAAGATCATGTTGGCTGACCTTATCGTCTCGCCGCGCAAGCT 575
QY 570 GCTTATGATGATCCATGGGCTTACCTGCTTACGCTTCTTTTCGGCGCGCTCGATATTG 629
Db 576 CGCCCTCGAGCACGCGCTTCGAGACCCCTCGGTTTCGGCGGTGGTGGCGGACACCTG 635
QY 630 GSAACCCGAAAAGATATCTACTGGGGTGACGA 662
Db 636 GGAAGCCGATGATGCTCGCTACTACTGGGGCCGA 668

RESULT 13
AW180683 573 bp mRNA linear EST 17-NOV-1999
LOCUS Mga0831f Mga Library Mycosphaerella graminicola cDNA clone Mga0831
DEFINITION 5', similar to catalase-peroxidase, mRNA sequence.
ACCESSION AW180683
VERSION AW180683.1 GI:6447878
KEYWORDS EST.
SOURCE Mycosphaerella graminicola
ORGANISM Mycosphaerella graminicola
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes et
Chaetothromycetes incertae sedis; Mycosphaerellaceae;
Mycosphaerella.
REFERENCE 1 (bases 1 to 573)
AUTHORS Keon,J.P.R., Bailey,A.M. and Hargreaves,J.A.
TITLE A group of expressed cDNA sequences from the wheat fungal leaf
blotch pathogen, Mycosphaerella graminicola (Septoria tritici)
JOURNAL Fungal Genet. Biol. 29 (2), 118-133 (2000)
MEDLINE 20374020
PUBMED 10919380
COMMENT Contact: Hargreaves JA
Cell Biology Department
IACR-Long Ashton Research Station
Long Ashton, Bristol, BS41 9AF, UK
Tel: +44(0)1275 392181
Fax: +44(0)1275 394281
Email: john.hargreaves@bbsrc.ac.uk
Insert length: 2300 Std Error: 0.00
Seq primer: M13 reverse.
Location/Qualifiers

```

BASE COUNT

Db 620 CGCCCTCGAGCACAGCGGCTTCGAGACCCCTCGGTTTCGCCGGTGTGTCGCCGCGACACCTG 679
QY 630 GGAACCCGAAAGATATCTAC 651
Db 680 NGAAGCCGATGATCCGCTAC 701

RESULT 15

BU062489/c
LOCUS 712 bp mRNA linear EST 26-AUG-2002
DEFINITION Fgr_1_N19_T7 Nitrogen-starved mycelia Gibberella zeae cDNA, mRNA
sequence.

ACCESSION BU062489

VERSION BU062489.1 GI:22502778

KEYWORDS EST.

SOURCE Gibberella zeae

ORGANISM Gibberella zeae

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.

REFERENCE

AUTHORS Trail, F., Xu, J.-R., San Miguel, P., Halgren, R.G. and Kistler, H.C.

TITLE Analysis of expressed sequence tags from Gibberella zeae (anamorph

Fusarium graminearum)

JOURNAL Fungal Genet. Biol. 38 (2), 187-197 (2003)

MEDLINE 22508120

PUBMED 12620255

COMMENT

Contact: Frances Trail

Department of Plant Biology

Michigan State University

East Lansing, MI 48824, USA

Tel: 517 432 2939

Fax: 517 353 1926

Email: trail@msu.edu

Plate: 1 row: N column: 19.

Location/Qualifiers

1..712

/organism="Gibberella zeae"

/mol_type="mRNA"

/strain="NRRL 31084"

/db_xref="taxon:5518"

/clone_lib="Nitrogen-starved mycelia"

/note="Vector: Uni-ZapII; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 152 a 184 c 211 g 164 t 1 others

FEATURES

source

Query Match 8.9%; Score 201.8; DB 13; Length 712;
Best Local Similarity 60.0%; Pred. No. 5.2e-44;
Matches 354; Conservative 0; Mismatches 233; Indels 3; Gaps 1;

QY 1665 GTGATCGTTCTGGCGGTAGCGTAGGATCGAGAAAGCGGCGGAGACGAGGTACGAT 1725

Db 706 GGGATCGTCTTGTGCTGCTGCTGCGTCGAGAGGCGGCGCAAGGATCGCGGTGTNAAC 647

QY 1726 GTGCGGTTCCCTTCTGAAAGCGGTGCGGATGCGACCGCGGAGATGACCGACGAGAC 1785

Db 646 GTCACCGTGCCCTTCTACCTCTGCGCGTACCGATGCCACCGGACCGACGACGCTCCAG 587

QY 1786 TCCTTCGACCGCTGGAGCGGCTGCGCGATGCTTCGCAACTGCGGAGAAAGAGTAT 1845

Db 586 TCTGTGCTCAGCTCGAGCGCGCGCGGAGGTTCCGGAATACGGCGCTCCACAGAC 527

QY 1846 GTGGTGAGCGCGGAGAGATGCTGTGGATGCTGCGAGCTGATGGGCTTAACGGCGCG 1905

Db 526 CGTGTCAAGCTCGAGCACATGCTCATCGACCGCGCCAGCTTCTCAGCGTCCOC 467

QY 1906 GAATGACCGTCTGCTGGCGGCTATGCGGCTACTGGGCGACCACTATGTTGGCACCAA 1965

Db 466 GAGATCACCGCTCTGCTGCGGTGCTTCTGTCGCTCAACGCCAAGTGGGACGATCATCC 407

QY 1966 CACGGCGTATTCACCGCATTTGAAGGCCAGTGTGACCAAGCTTTTTTGTGAACCTGACC 2025

Db 406 CACGGTATCTCACCAAGCGCCCTGGTGTCTGTCCAGACTTCTTCTGTCACCTCTCTC 347

Search completed: October 8, 2003, 15:29:18
Job time : 4695.37 secs

QY 2026 GATATGGGGAACAGCTGGAAGCCGCTAGGTAGCAACGCC---TAGGAATCCGACCGC 2082
Db 346 GACATCAGCACCGAGTGAAGCCCTCGGCAACGCCGACCTGTCGAGGGTATTGACCGC 287
QY 2083 AAGACCGGTGCGGTGAAGTGGACCGCCTCGCGGTGGATCTGTTGTTTCCAACTCG 2142
Db 286 AAGAGCGGCAACAAGAGTGGACTGTACCCGCTGCGATCTAGTCTTTGGATCACACGCT 227
QY 2143 CTACTGCGCTCTTACGCGAAGTGTACGCCAGGACGATTAACGGGAGAGAAGTTCGTACA 2202
Db 226 GAGCTCCGCGCGACAGCTGAGACATATGCTGAAGCTGGCGGTAGCGAGAAGCTGTTCCG 167
QY 2203 GACTTCGTGCGCGCTGGACCAAGTGTAGAACGCGGACCGCTTTCGACGT 2252
Db 166 GACTTAGTGTGCTGCGTGGACCAAGTGTAGAACCTTGACCGATTGATGT 117

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 7, 2003, 19:31:22 ; Search time 60.8231 Seconds
(without alignments)
1965.061 Million cell updates/sec

Title: US-09-884-889-6
Perfect score: 4058
Sequence: 1 MNASADLHSLQRCRAF.....RDFVAANTKVMNDRFDVAS 753

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4058	100.0	753	19 AAW33809	Alcaligenes (Deley)
2	4058	100.0	753	23 ABG33040	Alcaligenes catala
3	2820	69.5	723	23 ABG78212	Amino acid sequen
4	2290.5	56.4	745	19 AAW33810	Microscilla furves
5	2044.5	50.4	740	18 AAW28596	Mycobacterium tube
6	2044.5	50.4	740	20 AAW95398	M. tuberculosis ca
7	2041.5	50.3	740	19 AAW31343	Mycobacterium tube
8	2041.5	50.3	740	20 AAW95399	M. tuberculosis ca
9	1890	46.6	735	14 AAR43670	M. tuberculosis H37

10	1890	46.6	735	20	AAW78363	M.tuberculosis kat
11	1547	38.1	731	11	AAR08205	Heat resistant per
12	1050	25.9	533	22	ABG25057	Novel human diago
13	784.5	19.3	273	22	ABG08823	Novel human diago
14	519	12.8	1550	22	ABG09297	Novel human diago
15	398	9.8	201	22	ABG08828	Novel human diago
16	272.5	6.7	291	22	ABH03661	Hordeum vulgare pe
17	270	6.7	246	23	ABH92257	Herbicidally activ
18	269	6.6	354	21	AAG44544	Arabidopsis thalia
19	269	6.6	426	21	AAG44543	Arabidopsis thalia
20	269	6.6	426	23	ABH91634	Herbicidally activ
21	269	6.6	452	21	AAG44542	Arabidopsis thalia
22	267	6.6	250	21	AAG09702	Arabidopsis thalia
23	267	6.6	250	21	AAG16805	Arabidopsis thalia
24	267	6.6	250	21	AAG26453	Arabidopsis thalia
25	267	6.6	250	21	AAG43523	Arabidopsis thalia
26	267	6.6	250	21	AAG45582	Arabidopsis thalia
27	267	6.6	250	21	AAG45585	Arabidopsis thalia
28	267	6.6	250	21	AAG45589	Arabidopsis thalia
29	267	6.6	250	23	ABH90903	Herbicidally activ
30	267	6.6	254	21	AAG43522	Arabidopsis thalia
31	267	6.6	297	21	AAG26452	Arabidopsis thalia
32	267	6.6	297	21	AAG45588	Arabidopsis thalia
33	265	6.5	215	21	AAG09703	Arabidopsis thalia
34	265	6.5	215	21	AAG16806	Arabidopsis thalia
35	265	6.5	215	21	AAG26454	Arabidopsis thalia
36	265	6.5	215	21	AAG43524	Arabidopsis thalia
37	265	6.5	215	21	AAG45583	Arabidopsis thalia
38	265	6.5	215	21	AAG45586	Arabidopsis thalia
39	265	6.5	215	21	AAG45590	Arabidopsis thalia
40	260.5	6.4	347	21	AAG24604	Arabidopsis thalia
41	260.5	6.4	348	21	AAG05527	Arabidopsis thalia
42	260.5	6.4	353	21	AAG05526	Arabidopsis thalia
43	260.5	6.4	353	21	AAG24603	Arabidopsis thalia
44	260	6.4	279	23	ABH93261	Herbicidally activ
45	259.5	6.4	348	21	AAG52697	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAW33809
ID AAW33809 standard; Protein; 753 AA.
XX AC AAW33809;
XX DT 18-NOV-1998 (first entry)
XX DE Alcaligenes (Deleya) aquamarinus catalase-64CA2.
XX KW Catalase; epoxidation; hydroxylation; biosensor; paper bleaching;
XX KW pasteurisation.
XX OS Alcaligenes aquamarinus.
XX XX
XX PN WO9800526-A1.
XX PD 08-JAN-1998.
XX PF 03-JUL-1997; 97WO-US16513.
XX PR 03-JUL-1996; 96US-0674887.
XX XX
XX PA (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX PI Adhikary RS, Robertson DE, Sanyal I;
XX DR WPI; 1998-086953/08.
XX DR N-PSDB; AAV06554.
XX PT New bacterial catalases, related nucleic acid vectors and
PT transformed cells - used as oxidising agents and for detecting or

PT destroying hydrogen peroxide, e.g. in biosensors
PS Claim 1; Fig 1; 35pp; English.
XX

CC The present sequence is of the Alcaligenes aquamarinus catalase-6ACA2.
CC Catalase-6ACA2 may be used to catalyse oxidation reactions such as
CC epoxidation or hydroxylation. The enzyme can also be used to detect or
CC destroy hydrogen peroxide, e.g. in connection with glyoxylic acid
CC production, biosensors, contact lens cleaning, pulp/paper bleaching and
CC pasteurisation of dairy products. Antibodies raised against
CC catalase-6ACA2 can be used to screen libraries for detection and
CC purification of cells containing the enzyme. Fragments of the gene
CC encoding Catalase-6ACA2 can be used to identify related sequences.
XX

SQ Sequence 753 AA;

Query Match 100.0%; Score 4058; DB 19; Length 753;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 753; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MNNASADDLHSSLQORCRAFPVLPSPRHAIIRERAMSGKCPVMEGGNTSTGTSNKDWPE 60
Db 1 MNNASADDLHSSLQORCRAFPVLPSPRHAIIRERAMSGKCPVMEGGNTSTGTSNKDWPE 60
Qy 61 GLNLIDLHQDRKSDPMDPDFNRYEEVRKLDKDDVHALMTDSQEWMPADMGHYGGL 120
Db 61 GLNLIDLHQDRKSDPMDPDFNRYEEVRKLDKDDVHALMTDSQEWMPADMGHYGGL 120
Qy 121 MIRMAHSAGTYRTADRGGGGTGSGRFAPLNSPDVSLDKARLLWPIKPKYGNKISW 180
Db 121 MIRMAHSAGTYRTADRGGGGTGSGRFAPLNSPDVSLDKARLLWPIKPKYGNKISW 180
Qy 181 ADLMLAGTVAYESMGLPAYGFSGRVDIWEPRDIYWGDEKEWLAPSDERYGVNKPET 240
Db 181 ADLMLAGTVAYESMGLPAYGFSGRVDIWEPRDIYWGDEKEWLAPSDERYGVNKPET 240
Qy 241 MENPLAAVQMLIYVNEGVNGHPDURTAAQOVLTFARMANDEKTAALTAGGHTVNC 300
Db 241 MENPLAAVQMLIYVNEGVNGHPDURTAAQOVLTFARMANDEKTAALTAGGHTVNC 300
Qy 301 HGNGNASALADPKASDVENGGLGNPNMGKASNAVTSIGEGAWTNPTRKFGMGYFDL 360
Db 301 HGNGNASALADPKASDVENGGLGNPNMGKASNAVTSIGEGAWTNPTRKFGMGYFDL 360
Qy 361 LFQYNWELKSPAGAHWEPIIDIKENKPVDSADSPSRHNPIMTDADMAIKVNPYRAIC 420
Db 361 LFQYNWELKSPAGAHWEPIIDIKENKPVDSADSPSRHNPIMTDADMAIKVNPYRAIC 420
Qy 421 EKFMADPEYFKTKAKAFKLTDRDLGPKSRYIGPEVPAEDLIWQDPIPAAGNTDYCEEV 480
Db 421 EKFMADPEYFKTKAKAFKLTDRDLGPKSRYIGPEVPAEDLIWQDPIPAAGNTDYCEEV 480
Qy 481 KQKIAQSGLSISEMVTASDARSYRSGDMRGANGARIRLAPONEWQNEPERLAKVLS 540
Db 481 KQKIAQSGLSISEMVTASDARSYRSGDMRGANGARIRLAPONEWQNEPERLAKVLS 540
Qy 541 VYEQISADTGASTIADIVLAGSVIEKAAGAAGYDVRVPLFKRGDATAEMTADSFAPL 600
Db 541 VYEQISADTGASTIADIVLAGSVIEKAAGAAGYDVRVPLFKRGDATAEMTADSFAPL 600
Qy 601 EPLADGFRNQKKEYVYVVPKPEMLLDRAQLMGLTGPEMTVLGGMRVLTGNYGGTKHGVT 660
Db 601 EPLADGFRNQKKEYVYVVPKPEMLLDRAQLMGLTGPEMTVLGGMRVLTGNYGGTKHGVT 660
Qy 661 DCEQLNDFVNLTDGNSKWPVGSNAYETDRDKTGAVKWTASRVDLVFGSNSLLRSYA 720
Db 661 DCEQLNDFVNLTDGNSKWPVGSNAYETDRDKTGAVKWTASRVDLVFGSNSLLRSYA 720
Qy 721 EVTAQDDNGEKFVRDFFVAANTKVMNADRFDVAS 753
Db 721 EVTAQDDNGEKFVRDFFVAANTKVMNADRFDVAS 753

RESULT 2
ABG33040

ID ABG33040 standard; Protein; 753 AA.

XX AC ABG33040;

XX DT 22-NOV-2002 (first entry)

XX DE Alcaligenes catalase protein 6ACA2.

XX KW Gene; ss; marine bacteria; catalase; hydrogen peroxide; enzyme.

XX OS Alcaligenes aquamarinus.

XX PN US2002102680-A1.

XX PD 01-AUG-2002.

XX PF 19-JUN-2001; 2001US-0884889.

XX PR 03-JUL-1996; 96US-0674887.

XX PR 16-OCT-1997; 97US-0951844.

XX PR 05-OCT-1999; 99US-0412347.

XX PA (ROBE/) ROBERTSON D E.

XX PA (SANY/) SANYAL I.

XX PA (ADHI/) ADHIKARI R.

XX PI Robertson DE, Sanyal I, Adhikari R;

XX DR WPI; 2002-690613/74.

XX DR N-PSDB; ABS53944.

XX PT New purified catalase polypeptide useful for catalyzing the breakdown

XX PT of hydrogen peroxide, for modifying small molecules, and for generating

XX PT antibodies which bind to the polypeptide

XX PS Claim 22; Fig 5; 44pp; English.

XX CC This invention relates to the DNA and protein sequences of a novel

XX CC purified catalase polypeptide. The protein of the invention or its

XX CC homologue is useful for catalyzing the breakdown of hydrogen peroxide.

XX CC The protein sequence of the invention is useful for modifying small

XX CC molecules, by mixing the protein or its fragments with a small molecule

XX CC to produce a modified small molecule. The protein of the invention is

XX CC also useful for catalyzing biochemical reactions, for hydrolysing

XX CC glycosidic linkages and for generating antibodies which bind

XX CC specifically to the protein. The nucleic acid sequences of the

XX CC invention is useful as a probe to determine whether a biological sample,

XX CC such as a soil sample, contains an organism having the nucleic acid or

XX CC an organism from which the DNA was obtained, or for identifying an

XX CC isolating related nucleic acids. An antibody that binds the cellulase

XX CC protein of the invention is useful in immunaffinity chromatography

XX CC procedures to isolate or purify the protein or for detection of protein

XX CC expression in a biological sample. The cellulase protein of the

XX CC invention is heat stable, is heat resistant, and is able to

XX CC renature and regain activity after exposure to temperatures of from

XX CC about 60 to 105 degrees. The present sequence represents the Alcaligenes

XX CC aquamarinus catalase protein of the invention.

XX SQ Sequence 753 AA;

Query Match 100.0%; Score 4058; DB 23; Length 753;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 753; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MNNASADDLHSSLQORCRAFPVLPSPRHAIIRERAMSGKCPVMEGGNTSTGTSNKDWPE 60
Db 1 MNNASADDLHSSLQORCRAFPVLPSPRHAIIRERAMSGKCPVMEGGNTSTGTSNKDWPE 60
Qy 61 GLNLIDLHQDRKSDPMDPDFNRYEEVRKLDKDDVHALMTDSQEWMPADMGHYGGL 120
Db 61 GLNLIDLHQDRKSDPMDPDFNRYEEVRKLDKDDVHALMTDSQEWMPADMGHYGGL 120

QY 121 MIRMWHSAGTYRIADCGGGTGSORFAPLNSWPNVSLDKARLLWPIKKYGNKISW 180
 Db 121 MIRMWHSAGTYRIADCGGGTGSORFAPLNSWPNVSLDKARLLWPIKKYGNKISW 180
 QY 181 ADLMILAGVAVESMGLPAYGFSFGVVDLWEPEKDIYWGDEKWLAPSDRYGVNKPET 240
 Db 181 ADLMILAGVAVESMGLPAYGFSFGVVDLWEPEKDIYWGDEKWLAPSDRYGVNKPET 240
 QY 241 MENPLAAVOMGLIYVNPGEVNGHPDPLRTAQVLETFARMAMNDEKTAALTAGGHTVNC 300
 Db 241 MENPLAAVOMGLIYVNPGEVNGHPDPLRTAQVLETFARMAMNDEKTAALTAGGHTVNC 300
 QY 301 HGNHNASALAPDPKASDVENOGILGNGNPNMQKASNAVTSIGEGATWNTPTFDGIFYDL 360
 Db 301 HGNHNASALAPDPKASDVENOGILGNGNPNMQKASNAVTSIGEGATWNTPTFDGIFYDL 360
 QY 361 LFGYNWELKSPAGAHWHPIDIKKENKPVDAADSPISIRHNPIMTDADMAIKVNPYRALC 420
 Db 361 LFGYNWELKSPAGAHWHPIDIKKENKPVDAADSPISIRHNPIMTDADMAIKVNPYRALC 420
 QY 421 EKFMADPEYFKTKTAKANFKLHRDLGPKSRYIGPEVPAEDLIWQDPIAGNTDYCEEV 480
 Db 421 EKFMADPEYFKTKTAKANFKLHRDLGPKSRYIGPEVPAEDLIWQDPIAGNTDYCEEV 480
 QY 481 KOKIAQSGLSISEMSTANDSARTYRGSDMRGANGARIRLAPONWQNEPERLAKVLS 540
 Db 481 KOKIAQSGLSISEMSTANDSARTYRGSDMRGANGARIRLAPONWQNEPERLAKVLS 540
 QY 541 VYEOISADTGASTADIVILVAGSVIGIEKAAAGYDVRVFLKRGDGTAEWTDADSFAPL 600
 Db 541 VYEQISADTGASTADIVILVAGSVIGIEKAAAGYDVRVFLKRGDGTAEWTDADSFAPL 600
 QY 601 EPLADGFRNWQKEYVYVPEEMLLDRAQLMGLTGPEMTVLLGGMRYVLGTNYGTRHGVT 660
 Db 601 EPLADGFRNWQKEYVYVPEEMLLDRAQLMGLTGPEMTVLLGGMRYVLGTNYGTRHGVT 660
 QY 661 DCEGQLTNDFVNLTDGMSWPKVGSNAYEIRDRKTGAVKWTASRDVLDVFGSNLSRVA 720
 Db 661 DCEGQLTNDFVNLTDGMSWPKVGSNAYEIRDRKTGAVKWTASRDVLDVFGSNLSRVA 720
 QY 721 EYVAQDDNGEKFVRDFVAAWTKVMNADREFDVAS 753
 Db 721 EYVAQDDNGEKFVRDFVAAWTKVMNADREFDVAS 753
 RESULT 3
 ID ABB78212
 AC ABB78212 standard; Protein; 723 AA.
 XX ABB78212;
 DT 25-NOV-2002 (first entry)
 XX 25-NOV-2002 (first entry)
 DE Amino acid sequence of Agrobacterium tumefaciens catalase Kata gene.
 XX Catalase; Kata; reactive oxygen species; ROS; ROS-inducible promoter;
 KW H2O2 level; antioxidant.
 XX Agrobacterium tumefaciens.
 OS Agrobacterium tumefaciens.
 XX WO200263032-A1.
 PN 15-AUG-2002.
 XX 15-AUG-2002.
 XX 05-FEB-2002; 2002WO-5600018.
 XX 05-FEB-2002; 2001US-266657P.
 PR (PANS/) PAN S Q.
 XX (PANS/) PAN S Q.
 XX Pan SQ;
 PI Pan SQ;
 XX

WPI: 2002-643421/69.
 N-PSDB; ABQ78599.

Determining the ability of a compound to remove a reactive oxygen species (ROS), useful in pharmaceutical and nutraceutical fields and measuring activities of antioxidants, comprises providing a cell having a ROS-inducible promoter -

Disclosure; Page 63-67; 70pp; English.

The present sequence represents Agrobacterium tumefaciens catalase (Kata). Kata has a promoter which is induced by a reactive oxygen species (ROS). The kata gene can be used in the method of the invention. The specification describes a method for determining the ability of a compound to remove a ROS. The method comprises providing a cell having a ROS-inducible promoter operably linked to a reporter gene that is heterologous to the promoter, and exposing the cell to a compound potentially able to remove the ROS. The method is useful in determining the ability of a compound or gene product to remove a ROS and determining the level of the ROS H2O2 inside a cell, to properly assess and label antioxidant products for use as food supplements, natural products or drugs. The method may also be used in selecting for new or improved ROS scavengers.

Sequence 723 AA;

Query Match 69.5%; Score 2820; DB 23; Length 723;
 Best Local Similarity 71.0%; Pred. No. 6.8e-236;
 Matches 508; Conservative 86; Mismatches 121; Indels 0; Gaps 0;

QY 37 SKCPVHCGNTSTGTSTNKNWPEGLNLDILHQDRKSDPMDPDFWYREVRLDPAUK 96
 Db 8 AKCPVHCGNTSTGTSTNKNWPEGLNLDILHQDRKSDPMDPDFWYREVRLDPAUK 96
 QY 97 KDVALMTDSQEWPPADWGHYGLMIRMAWHSAGTYRIADCGGGTGSORFAPLNSWPD 156
 Db 68 ADRLALMTDSQEWPPADWGHYGLMIRMAWHSAGTYRIADCGGGTGSORFAPLNSWPD 156
 QY 157 NVSLDKARLLWPIKKYGNKISWADLMILAGTVAVESMGLPAYGFSFGVVDLWEPEKDI 216
 Db 128 NVNTDKGRLLWPIKKYGNKISWADLMILAGTVAVESMGLPAYGFSFGVVDLWEPEKDI 216
 QY 217 YWDEKEWLAPSDRYGVNKPETMENPLAAVOMGLIYVNPGEVNGHPDPLRTAQVLET 276
 Db 188 YWDEKEWLAPSDRYGVNKPETMENPLAAVOMGLIYVNPGEVNGHPDPLRTAQVLET 276
 QY 277 FARMAMNDEKTAALTAGGHTVNGCHGNASALAPDPKASDVENOGILGNGNPNMQKAS 336
 Db 248 FARMAMNDEKTAALTAGGHTVNGCHGNASALAPDPKASDVENOGILGNGNPNMQKAS 336
 QY 337 AVTSGIEGAWTNTPTKEDMGYFDLLFGYNWELKSPAGAHWHPIDIKKENKPVDAADSP 396
 Db 308 TVVSGIEGAWTNTPTKEDMGYFDLLFGYNWELKSPAGAHWHPIDIKKENKPVDAADSP 396
 QY 397 IRHNPIMTDADMAIKVNPYRALCERFADPEYFKTKTAKANFKLHRDLGPKSRYIGPE 456
 Db 368 IRTPMTADMAIKVNPYRALCERFADPEYFKTKTAKANFKLHRDLGPKSRYIGPE 456
 QY 457 VPAEDLIWQDPIAGNTDYCEEVVKOKIAQSGLSISEMSTANDSARTYRGSDMRGANG 516
 Db 428 VPAEDLIWQDPIAGNTDYCEEVVKOKIAQSGLSISEMSTANDSARTYRGSDMRGANG 516
 QY 517 ARIRLAPONWQNEPERLAKVLSYVEIQISADTGASTADIVILVAGSVIGIEKAAAGYDV 576
 Db 488 ARIRLAPONWQNEPERLAKVLSYVEIQISADTGASTADIVILVAGSVIGIEKAAAGYDV 576
 QY 577 RVPFLKRGDGTAEWTDADSFAPLEPLADGFRNWQKEYVYVPEEMLLDRAQLMGLTGP 636
 Db 548 AVPFAAGRGDASAEQTDADSFAPLEPLADGFRNWQKEYVYVPEEMLLDRAQLMGLTGP 636
 QY 637 MTVLLGGMRYVLGTNYGTRHGVTDCGQLTNDFVNLTDGMSWPKVGSNAYEIRDRKT 696
 Db 608 LTVLLGGMRYVLGTNYGTRHGVTDCGQLTNDFVNLTDGMSWPKVGSNAYEIRDRKT 696

```

QY 697 GAVKWTASRVLDVFGSNLSRLSYAEVYAQDDNGEKFVRDFVAAWTKVMNADRFV 751
DB 668 GAARYSATRVLDVIGSNLSILRAYAEVYAQDDNREKFAADFATAAWTKVMNADRFDL 722

RESULT 4
AAW33810
ID AAW33810 standard; Protein: 745 AA.
XX
AC AAW33810;
XX
DT 18-NOV-1998 (first entry)
XX
DE Microscilla furvescens catalase-53Cal.
XX
KW Catalase; epoxidation; hydroxylation; biosensor; paper bleaching;
KW pasteurisation.
XX
OS Microscilla furvescens.
XX
PN WO9800526-A1.
XX
PD 08-JAN-1998.
XX
PF 03-JUL-1997; 97WO-US16513.
XX
PR 03-JUL-1996; 96US-0674887.
XX
PA (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX
PI Adhikary RS, Robertson DE, Sanyal I;
XX
DR WPI; 1998-086953/08.
XX
DR N-PSDB; AAV06555.
XX
PT New bacterial catalases, related nucleic acid vectors and
PT transformed cells - used as oxidising agents and for detecting or
PT destroying hydrogen peroxide, e.g. in biosensors
XX
PS Claim 1; Fig 2; 35pp; English.
XX
CC The present sequence is of the Microscilla furvescens catalase-53Cal.
CC Catalase-53Cal may be used to catalyse oxidation reactions such as
CC epoxidation or hydroxylation. The enzyme can also be used to detect or
CC destroy hydrogen peroxide, e.g. in connection with glyoxylic acid
CC production, biosensors, contact lens cleaning, pulp/paper bleaching and
CC pasteurisation of dairy products. Antibodies raised against
CC catalase-53Cal can be used to screen libraries for detection and
CC purification of cells containing the enzyme. Fragments of the gene
CC encoding Catalase-53Cal can be used to identify related sequences.
XX
SQ Sequence 745 AA;

Query Match 56.48; Score 2290.5; DB 19; Length 745;
Best Local Similarity 59.18; Pred. No. 7.8e-190;
Matches 433; Conservative 92; Mismatches 183; Indels 25; Gaps 8;

QY 38 GKCPVMHGG---NTSTGTSNKRDMWPEGLINDILHQDQKSDPMDPDFNYREEVRKLDFDA 94
DB 18 GKCPFTGSLKQSAGGSKRDWPNMLNLGILRQHSLSLSDPNDDPDYAEFEKLLDLAA 77
QY 95 LKDVHALMTDSQEWPMADGHYGGGLIRMAHWSAGTYRIADGRGGGTGSGRQAPLNSW 154
DB 78 VKDLAALMTDSQDWPADYGHYGPFFIRMAHWSAGTYRICDGRGGGSGSRQAPLNSW 137
QY 155 PDNVSLDKARLLMPIKKYKNTKISWADLMILACTVAYESMGLPAYGFSFGRVDIWRPEK 214
DB 138 PDNALDLKARLLMPIKQYGRKISWADLMILGNVALETMGFKTFGAGRADVWPEE 197
QY 215 DIYWGDEKWLAPSDERY-GDVNKPETMENPLAIVQMGGLIYVNPGEVNGHDPDLRTAQV 273
DB 198 DVTWGAETEWL--GDKRYEGD----RELENPLGAVQMGGLIYVNPGEVNGKPGDPIAAARDI 251

```

```

QY 274 LETFARMAMDEKTAALTAGGTVGNCHGNCGNASA-LAPDPKASDVENQGLGWGNPNMOG 332
DB 252 RETFGRMANMDEETVALTAGGTFGKTHGADADAEKYVGREPAAAGIEMSLGNWNTYGTG 311
QY 333 KASNAVSGIGEGAWTNTPTKFDMGYFDLLFYNNMELKKSPAGAHWEPIDIKKENKPVDA 392
DB 312 HGADTITSGLEGAWTKTPTQWNNFFENLFCYEWELTKSPAGAYQWKPKDCAGAGTIPDA 371
QY 393 SDPSIRHNPIMTDADMAIKVNPTYRAICEKFWADPEYFKTFAKAWFKLTIRDLGPKSRY 452
DB 372 HDPKSHAPFMLTTDLALRMDPDYEKISRYYENPDEFADAFKAWFKLTHRDMGPKVRY 431
QY 453 IGPEVPAEDLIWQDPIPAAGNTDYCE---EYVVKQIAQSGLSISEMSTVSTANDSARTYRGS 508
DB 432 LGPEVPQEDLIWQDPIPDVSHPLVDENDIEGLKAKILLESGLTVSELVSTAWASSTFRNS 491
QY 509 DMRGANGARIRLAPQNEWQNEPERLAKVLSVYEQISAD-----TGASIAADVIVLA 560
DB 492 DKRGANGARIRLAPQKDWVNNPQQLARVLKTLLEGIOEDFNQASQNSONKAVSLADLIVLA 551
QY 561 GSVGIEKAAKAGYDVRVPELKGKRGDATAETADSPAPLEPLADGFRNWKKEVYVYKPE 620
DB 552 GCAGVERAAKDAGHEVQVFPNPGRADATAEGTDEAFEALPEADGFRNYIKPEHKVSAE 611
QY 621 EMLLDRAQLMGLTGPEMTVLGGMRVLGTYNTGGTKHGVTTCDEGLQTNDFVNLTDGNS 680
DB 612 EMLVDRAQLLSLSAPEMTALVGGMRVLGTYNDGSHQGVFTNKPQLSNDFFVNLIDLNTK 671
QY 681 WK--PVGSNAYEIRDRKTGAVKWTASRVLDVFGSNLSRLSYAEVYAQDDNGEKFVRDFVA 738
DB 672 WRASDESKVPEGRDFTKEVKGSGTRVDLIFGNSSELRALAEVYGACDSEKFKVDPVK 731
QY 739 AWTVMNADRFV 751
DB 732 AWAKVMDLDRFDL 744

RESULT 5
AAW26596
ID AAW26596 standard; Protein: 740 AA.
XX
AC AAW26596;
XX
DT 25-MAR-2003 (updated)
DT 21-JAN-1998 (first entry)
XX
DE Mycobacterium tuberculosis wild-type catalase-peroxidase.
XX
KW Isoniazid resistance; restriction fragment length polymorphism;
KW RFLP; katG gene; catalase; peroxidase; multiple drug resistance;
KW isonicotinic acid hydrazide; INH; tuberculosis; antibiotic;
KW tuberculostatic.
XX
OS Mycobacterium tuberculosis.
XX
FH Key Location/Qualifiers
FT Misc-difference 264
FT /note= "mutation to Thr confers INH resistance"
FT Misc-difference 315
FT /note= "mutation to Thr confers INH resistance"
FT Misc-difference 337
FT /note= "mutation to Cys confers INH resistance"
FT Misc-difference 463
FT /note= "mutation to Leu confers INH resistance"
XX
US5658733-A.
PN
XX
XX 19-AUG-1997.
XX
XX 07-APR-1995; 95US-0418782.
XX
XX 07-APR-1995; 95US-0418782.

```


PR	18-APR-1994;	9405-0228662.	
XX	(MAYO-)	MAYO FOUNDATION.	
XX			
XX	Cockerill FR,	Kline BC, Uhl JR;	
XX			
XX	WPI; 1997-424226/39.		
DR	N-PSDB; AAT90400.		
XX			
PT	Determining susceptibility of Mycobacterium tuberculosis strains to		
XX	isoniazid - by detecting mutation(s) in the catalase-peroxidase gene,		
PT	katG		
XX			
XX	Example 2; Column 31-36; 38pp; English.		
XX			
CC	This polypeptide comprises a consensus of the Mycobacterium		
CC	tuberculosis wild-type catalase-peroxidase enzyme. Claimed		
CC	methods of rapidly identifying strains of M. tuberculosis which are		
CC	resistant to the tuberculostatic drug Isoniazid (INH) are based on		
CC	the discovery of 4 mutations in the wild-type katG gene sequence		
CC	(see AAT90400) that confer INH resistance and which coincidentally		
CC	result in the addition or deletion of restriction endonuclease		
CC	sites. Primers (see AAT90401-02) are provided for the PCR		
CC	amplification of a katG gene from a test M. tuberculosis strain,		
CC	and further primers (see AAT90403-06) for use in the restriction		
CC	fragment length polymorphism analysis of the amplified gene and		
CC	hence determination of the susceptibility to INH of the strain.		
CC	(Updated on 25-MAR-2003 to correct PF field.)		
XX			
XX	Sequence	740 AA;	
XX			
XX	Query Match	50.4%; Score 2044.5; DB 18; Length 740;	
XX	Best Local Similarity	54.3%; Pred. No. 1.9e-168;	
XX	Matches 406; Conservative 93; Mismatches 220; Indels 29; Gaps 13		
QY	24	VSPHRAIRER---AMSGCPVM-HGGNTSTGTSTNDDWPEGLNLIILHQDRKSDPMDP 79	
DB	1	VPEQHPPTETTTGAASGCGPVGHMKYPVEGGGNDWPNRLNLKVLHONPAVADPMGA 60	
QY	80	DENYREEYRKLDFOALKVDVHALMTDSQEWPPADWGHYGGIMTRMAWHAGTYRTADGRG 139	
DB	61	AFDYAAEYATIDVDALTREDIEVMTTSQPPWADYGHYGPLFRMAWHAGTYRTIDGRG 120	
QY	140	GGGTGSGRFAPLNSWPNVDLSDKARRLLWPIKKYGNKISWADMLTAGTVAYESMGLPA 199	
DB	121	GAGGQGNORFAPLNSWPNVDNASLDKARRLLWPIKKYGNKLSWADLVFAGNCALESMGFT 180	
QY	200	YGFSRGYVDIWEPEKDIYWGDEKWLAPSDERYGDVKNKPTMENPLAAVQMGILYVNPES 259	
DB	181	FGFGFRGVQDNEPD-EVTVGKEATWL--GDERY---SGKRDLNPLAAVQMGILYVNPES 234	
QY	260	VNGHPDLPLTAQOVLTEPFARMAMNDEKTAALTAGGHTVGNCHGNASALADPPKASDYE 319	
DB	235	PNGNPDPMAAAVADIRETPRPMAMNDVETAAIIVGGHTFGKTHGAGPADLVGPEPEAAPE 294	
QY	320	NOGLGWGNPNWOGKASNAVTSIGEATNPTTFDMDGYFOLLFGYWNWELKKSPAGAHWE 379	
DB	295	QMLGKWKSSYGTGTGKDAITSGIEVVMVNTPTKWDNSFLFLEILYGEWELTKSPAGAWQT 354	
QY	380	PID-IKKENKPVDSDDPSIRNPTMDADMAIKVNPYTRAIACEKEMADPEYFKKTFAKAW 438	
DB	355	AKDGAGAGTIPDPFGGPG--RSPMTLATDSLVRDPIYERITRWLEHPEELADEFAKAW 412	
QY	439	FKLTHRLDLPKRSRYIGPEVPAEDLIWDDPTIPAGNTDYCEBV----VKQIATGSLGISBM 494	
DB	413	YKLIHRDGMGPVARYLGLPVKQTLMDQPPVPVNSHDLVGEAEATSKSQIRASGLTVSOL 472	
QY	495	VSTANDSARTYNGSDMRGGANGARIRLAPQNEWOGNEPE-RLAKVLSVYEQI-----SAD 548	
DB	473	VSTAWAAASSFRGSDKRGANGRIRLQPOVGWEVNDPDGDLRKVIRTLEEIOESFNSAA 532	
QY	549	TG---ASTADVILVAGSVGEKAAKAGYDVRVPFLKGRGDGTAEMTDDADSPAPLEPLAD 605	

QY 24 VSPRRRAIRER-----AMSGKCPVM-HGGNTSTGTSTNKNWPEGLNLDILHQDRKSDPMDP 79
 DB 1 VPEQHPITETTTGAASNGCPVGHMKYPVEGCGNQDWPNRNLKVLHQNPAVADPMGA 60
 QY 80 DFNREVRKLDALDKVDHALMTDSQEWPPADWGHYGLMIRMAHWSAGTYRIADGRG 139
 DB 61 AFDAAEVATIDVALTRDIEEYMTTSSQWPPADYGHYGLFIRMAHWSAGTYRIADGRG 120
 QY 140 GGGTSGREAPLNSWPDNSLDKARLLWPKIKKYNKISWADLMILAGTVAYESMGLPA 199
 DB 121 GAGGGMORAPLNSWPDNSLDKARLLWPKIKKYNKISWADLMILAGTVAYESMGLPA 180
 QY 200 YGSGFRGVDIWEPEKDIYNGDEKELAPSDERYGDVKNPETMENPLAAVQMGIIYVNEG 259
 DB 181 FGGGFRGVDQWEPD-EVYWGKEATWL--GDERY---SGKRDLENPLAAVQMGIIYVNEG 234
 QY 260 VNGHPDPLTAQOVLTFARMAMNDEKTAALTAGGHTVGNCHGNASALADPPKASDVE 319
 DB 235 PNGNPDPMAAAADIRETFRMAMNDVETAAIYVGSHTFGKTHGAGPADLVGPEEAPLE 294
 QY 320 NOGLGNGNPMCKASNAVTSIGEGAWTTPKTFDMGYFDLLFGYNKELKSPAGAAHWE 379
 DB 295 QMGLGWKSSYGVTGKDATSGIEVVTWTPKWDNSFLEILGYEWELTKSPAGAWQT 354
 QY 380 PID-IKKENKPVASDPSIRHNPIMTDADMAIKVNPYRAICEKFMADPEYFKKTFPAKAW 438
 DB 355 AKDGAGAGTIPDPFGGPG--RSPTMLATDLSLRVDPYERITRRWLEHPEELADEFAKAW 412
 QY 439 FKLTHDLGPKSKRYIGPEVPAEDLIWODPIAGNTDYCEEV----VKQIAQSGLSISEM 494
 DB 413 YKLHDMGPVARYLGLPVKOTLLWQDVPVAVSHDLVGEAEIASLKQSIRASGLTVSQL 472
 QY 495 VSTANDARTYSGDMRGANGARIRLAPQNEWQGNPE-RLAKVLSVYEQI-----SAD 548
 DB 473 VSTAWAASFRGSDRGANGRIRLQPVQWGVNDPDGLKRVITLLEEQESNSAA 532
 QY 549 TG---ASIAADVIVLAGSVGIEKAAAGYDVRVFLKRGDGTATAEMTDADSAPLEPLAD 605
 DB 533 PGNIKVSFADLVVIGCAIEAKAAAGHNITVPFTPGRTDASQEQTDVESEFAVLEPKAD 592
 QY 606 GFRRMOKKEVVKPEMLDRAQLMGLTGPMTVLLGGMRVLGTNYGGTKHGVFTDCEG 665
 DB 593 GFRRYLGKGNPLPAETMYLDDKANLLTSAPEMTVLVGGURVLGANTYKRLPLGVFTSEAS 652
 QY 666 LTNDFFVNLTDGNSWK--PVGSNAYEIRDRKTGAVKWTASRVLDVFGSNLSLRSAEYV 723
 DB 653 LTNDFFVNLDMGITWEPSPADGTYQGD-GSGKVKWTGSRVDLVFGSNLSLRSAEYV 711
 QY 724 AQDDNGEKFVDFVAAWTKVMNADRPDV 751
 DB 712 GADDAQPKFQVDFVAAWTKVMNADRPDV 739
 RESULT 7
 AAW31343
 ID AAW31343 standard; Protein; 740 AA.
 AC AAW31343;
 XX AAW31343;
 DT 20-MAR-1998 (first entry)
 DE Mycobacterium tuberculosis katG protein consensus sequence.
 XX Tuberculosis; katG gene; isoniazid resistance; INH;
 KW Isonicotinic acid hydrazide; consensus.
 XX Mycobacterium tuberculosis.
 OS
 XX US5688639-A.
 PN
 XX 18-NOV-1997.

PF 18-APR-1994; 94US-0228562.
 XX 18-APR-1994; 94US-0228562.
 XX (MAYO-) MAYO FOUNDATION.
 XX Cockerill FR, Kline BC, Uhl JR;
 XX WPI; 1998-007975/01.
 PT Determination of isoniazid sensitivity of Mycobacterium tuberculosis
 PT strains - by restriction length polymorphism analysis of katG gene
 XX Example 2; Column 15-20; 18pp; English.
 PS This protein sequence represents the consensus amino acid sequence of
 CC the katG gene from Mycobacterium tuberculosis strains H37Rv MC, ATCC
 CC 25618, ATCC 27994, G6108, AAR35827, L6627-92, AAL11150, CC
 CC AAL24204, AAL33308, AAL16980, L1781, TMC 306, AAL10373 and AAL23361. This
 CC sequence is which is used in a novel method to rapidly identify strains
 CC of M. tuberculosis which are resistant to isoniazid (INH, isonicotinic
 CC acid hydrazide). The method involves the use of restriction fragment
 CC length polymorphism (RFLP) analysis to determine if a NciI-MspI
 CC restriction site is absent in the DNA of the strain at the codon
 CC corresponding to codon 463 of this katG gene consensus sequence. The
 CC absence of the site indicates an INH-resistant strain.
 XX Sequence 740 AA;
 SQ
 Query Match 50.3%; Score 2041.5; DB 19; Length 740;
 Best Local Similarity 54.4%; Pred. No. 3.5e-168;
 Matches 405; Conservative 92; Mismatches 218; Indels 29; Gaps 13;
 QY 28 HRAIRER---AMSGKCPVM-HGGNTSTGTSTNKNWPEGLNLDILHQDRKSDPMDP 83
 DB 5 HPITETTTGAASNGCPVGHMKYPVEGCGNQDWPNRNLKVLHQNPAVADPMGA 64
 QY 84 REEVRLKDLDAKDKVHALMTDSQEWPPADWGHYGLMIRMAHWSAGTYRIADGRGGGT 143
 DB 65 AAEVATIDVALTRDIEEYMTTSSQWPPADYGHYGLFIRMAHWSAGTYRIADGRGGGT 124
 QY 144 GSGREAPLNSWPDNSLDKARLLWPKIKKYNKISWADLMILAGTVAYESMGLPA 203
 DB 125 GMRGAPLNSWPDNSLDKARLLWPKIKKYNKISWADLMILAGTVAYESMGLPA 184
 QY 204 FGRVDIWEPEKDIYNGDEKELAPSDERYGDVKNPETMENPLAAVQMGIIYVNEG 263
 DB 185 FGRVDQWEPD-EVYWGKEATWL--GDERY---SGKRDLENPLAAVQMGIIYVNEG 238
 QY 264 PDPLRTAQOVLTFARMAMNDEKTAALTAGGHTVGNCHGNASALADPPKASDVE 323
 DB 239 PDPMAAAADIRETFRMAMNDVETAAIYVGSHTFGKTHGAGPADLVGPEEAPLE 298
 QY 324 GSGNPNMCKASNAVTSIGEGAWTTPKTFDMGYFDLLFGYNWELKSPAGAHWE 382
 DB 299 GMRSSYGTGKDATSGIEVVTWTPKWDNSFLEILGYEWELTKSPAGAWQT 358
 QY 383 IKKENKPVASDPSIRHNPIMTDADMAIKVNPYRAICEKFMADPEYFKKTFPAK 442
 DB 359 AGAGTIPDPFGGPG--RSPTMLATDLSLRVDPYERITRRWLEHPEELADEFAK 416
 QY 443 HRDLGPKSKRYIGPEVPAEDLIWODPIAGNTDYCEEV----VKQIAQSGLSISEM 498
 DB 417 HRDMGPVARYLGLPVKOTLLWQDVPVAVSHDLVGEAEIASLKQSIRASGLTV 476
 QY 499 WDSARTYSGDMRGANGARIRLAPQNEWQGNPE-RLAKVLSVYEQI-----SAD 550
 DB 477 WAAASSFRGSDRGANGRIRLQPVQWGVNDPDGLKRVITLLEEQESNSA 536
 QY 551 -ASIAADVIVLAGSVGIEKAAAGYDVRVFLKRGDGTATAEMTDADSAPLE 609
 DB 537 KVSFADLVVIGCAIEAKAAAGHNITVPFTPGRTDASQEQTDVESEFAVLEPK 596

OS Mycobacterium tuberculosis (strain H37Rv).
 PN W09322454-A1.
 PD 11-NOV-1993.
 XX 30-APR-1993; 93WO-EP01063.
 XX 17-SEP-1992; 92FR-0011098.
 PR 30-APR-1992; 92US-0875940.
 PR 14-AUG-1992; 92US-0929206.
 PR 16-APR-1993; 93FR-0004545.
 XX (ASSI-) ASSISTANCE PUBLIQUE.
 PA (INSP) INST PASTEUR.
 PA (MEDI-) MEDICAL RES COUNCIL.
 PA (UYBE-) UNIV BERNE.
 PA (UYPA-) UNIV CURIE PARIS VI P & M.
 XX Bodmer T, Cole S, Heym B, Honore N, Telenti A;
 PI Young D, Zhang Y;
 XX WPI: 1993-368812/46.
 DR N-PSDB; AAQ51531.
 XX Rapid detection of antibiotic resistance in Mycobacteria - esp.
 PT isoniazid, rifampicin or streptomycin resistance in tuberculosis
 PT by detecting mutation in katG, rpoB or rpsL genes
 PS Example 2; Fig 8; 97pp; English.
 XX The katG gene of Mycobacterium tuberculosis was isolated on a KpnI
 CC fragment by shotgun cloning. The sequence contains one open reading
 CC frame with high coding probability; a 735 amino acid protein is
 CC encoded by the ORF having predicted mol.wt. 80029. This protein
 CC conforms to the structure of known bacterial catalase-peroxidases
 CC which consist of two modules (possibly the result of a duplication
 CC event), both showing homology to the yeast enzyme, fused to a unique
 CC N-terminal sequence of about 50 amino acid residues. Mutations in
 CC the katG gene can result in loss of INH-susceptibility, i.e.
 CC the formation of antibiotic resistant Mycobacterial strains
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX Sequence 735 AA;
 SQ
 Query Match 46.6%; Score 1890; DB 14; Length 735;
 Best Local Similarity 51.3%; Pred. No. 5.1e-155;
 Matches 380; Conservative 95; Mismatches 240; Indels 26; Gaps 12;
 QY 27 RHRAIRER---AMSGKPVW-HGNTSTGTSTKDWNPGLNLDLTHQODRKSDPMDPDFN 82
 DB 4 QHPPIITTTGAASNGCPVGHMKYPVEGGNQDWPENRLNLKVLHQNPAVADPMGAADF 63
 QY 83 YREVRKLDPAKDKVHALTDSQEWKPADMGHYGGLMIRMAHWSAGTYRIADRGGG 142
 DB 64 YAEVATSRDLATRLDEETVTSQPMWADYGHYGLFIRMAHAAAGTYRIHDGCGAG 123
 QY 143 TGSQRFAPLNSWPNVSLDKARLLNPIKKYGNKISWADLMILAGTVAYESMGLPAYF 202
 DB 124 GGNQRFAPLNSWPNASLDKARLLNPPVKYKKLWSADLIVPAGNRCARNWASRRSG 183
 QY 203 SPGRVDIWEPEKDIYWGDEKELAPSDERYGVNKPETMENPLAAVQMGLIYVPEGVNG 262
 DB 184 SASGVDQWETD-EVYWCKEATWL---GGDGY---SVSDLENPLAAVQMGLIYVPEANG 236
 QY 263 HPDPLRTAQOVLTEFARMANDEKTAALTAGGHTVGNCHGNASALAPPKASDVENQG 322
 DB 237 NPDPMAAAVDIRETFRMANNDVETAAALIVGCTFTGKTHGAGPADLVGPEEAPLEQMG 296
 QY 323 LGWGNPNMOGKASNAVTSIGTEGAWTNTPTKFDLMGYFDLLFGYNWELKKSPAGAHWEPI 382
 DB 297 LGWSSVGTGTGKDAITSGIEVWNTPTKWDNSFLBILGYEWELTKSPAGAWQYTA 356

QY 383 -IKKENKPVSDSPSIRHNPIMTDADMAIKVNPYTRAIKCEKFMADPEYFKKTFKAWFKL 441
 DB 357 GAGAGTIPDPGPGG--RSPMTLATDLRLVDPIYERITRWLEHPEPELADEFKAWYKL 414
 QY 442 THRLGPKSRVIGPEVAEDLIWQDPIPA-CNTDYCEEV--VKOKIAQSGLSISEMVSTA 498
 DB 415 IHRDMGPVARYLGLPLVPKQTLWQDPVPAVSTTSKQIASLQIRASGLTVSOLVSTA 474
 QY 499 WDSARTYRGSMDRGANGARIRLAPONEWQNEPERLAKVLSVVEQIS-----ADTGAS 552
 DB 475 WAAASSFRGSKRGANGRRIRLPQVGMWEVNDPDGSAQGHSHPEEIQESTRGRNKKVS 534
 QY 553 IADVIVLAGSVGIEKAAAGYDVRVFLKRGDATAEMTDADSFAPLEADGFRNWK 612
 DB 535 FADLVVLGGCAPLEKAAKAAGHNITVPTPGPHDASQEQTDVESFAVLEPKADGFRNVLG 594
 QY 613 KEYVVKPEEMLLDRAQLMGLTGPEMVLVLLGGMRVLTGNYGKIGVFTDCGGQLTNDFFV 672
 DB 595 KGNCRPSTSLDKANLLTUSAPEMTVLGGRLVIGANYKRLPLGLGVTEASESITNDFFV 654
 QY 673 NLTDGNSWK--PVGSNAYEIRDRKTGAVKWTASRVLDLVFGSNLLRSYAEVYAODDNGE 730
 DB 655 NLLDMGITWEPSPADDTYQKGD-GSGKVKWTGSRVDLVFGSNELRALVETVAPWTRQA 713
 QY 731 KFYRDFVAAMTKVMNADRFV 751
 DB 714 KFTVGFVAAMDKVMNLDREDFV 734
 RESULT 10
 AAW78363
 ID AAW78363 standard; Protein; 735 AA.
 AC AAW78363;
 XX 20-MAR-2003 (updated)
 DT 11-MAY-1999 (first entry)
 DE M.tuberculosis katG protein.
 XX katG; catalase; peroxidase; enzyme; isonicotinic acid hydrazide;
 KW isoniazid; INH; diagnosis; resistance.
 XX Mycobacterium tuberculosis.
 XX US5871912-A.
 PD 16-FEB-1999.
 XX 02-JUN-1995; 95US-0459499.
 PR 02-JUN-1995; 95US-0459499.
 PR 30-APR-1992; 92US-0875940.
 PR 14-AUG-1992; 92US-0929206.
 PR 11-MAR-1993; 93US-0029655.
 XX (INSP) INST PASTEUR.
 PA (MEDI-) MEDICAL RES COUNCIL.
 PA (UYPA-) UNIV PARIS VI.
 PI Cole ST, Heym B, Young DB, Zhang Y;
 XX WPI: 1999-166622/14.
 DR N-PSDB; AAX16963.
 XX New nucleic acid encoding catalase-peroxidase HPI - useful for rapid
 PT identification of isoniazid resistant Mycobacterium tuberculosis
 PT Disclosure; Fig 10A; 46pp; English.
 XX This sequence represents the catalase-peroxidase enzyme encoded by
 CC the Mycobacterium tuberculosis katG gene. The gene was isolated from
 CC a M. tuberculosis strain resistant to isonicotinic acid hydrazide

CC (isoniazid or INH). The sequence is useful in the diagnosis of
CC tuberculosis caused by INH-resistant *M. tuberculosis* and in the
CC identification of other INH-resistant strains.
CC (updated on 20-MAR-2003 to correct PA field.)

Sequence	735 AA;
SQ	

Query Match	46.6%	Score 1890;	DB 20;	Length 735;
Best Local Similarity	51.3%	Pred. No. 5.1e-155;		
Matches 380;	Conservative	95;	Mismatches 240;	Indels 26;
				Gaps 13;

Qy 27 RHRAIRER--AMSGKCPVM-HGGNTSTGTSTNKDWWPEGNLIDLHQQDRKSDPMDPDFN 82

Qy 83 YBEEVRKLDLAKKDVHALMTDSQEWNPADWGHYGGGLMIRMAHWSAGTYPIADCRGGGG 142

Db 64 YAAEVATSRDLALTRDIEEVMTTSQPNWPADYGHYGLPLFIRMAWHAAGYRIHDPGRGGAG 123

[illegible]

QY 203 SGRVDIWEPEXDIYWGDEKWLAPSDRRYGDYKNKPETMENPLAAVQMGLIYYNPEGVNG 262

00 00 104 263 Qv

Db 237 NPDMAAVIDIRETFRMAMDVETAALIVGGHTFGKTHGAGPADLVGPEEAPLEQMG 296

QY	323	LGWGNPAMQKASNAVTSIGIEGAWTNTPTKFDMGYFDLLFGYNWELKSPAGAAHWPED	388
DB	297	LGKWSYSGTGTGRDAITSGIEFVVTNTPTKWDNSPTITLVGYEKEITKSDAGACQVATAD	356

QY 383 -IKKENPVDASPSIRHNPIWTDADMAIKVNPTYRAICEKFMADPEYFKKTFKAKAWFKL 441

442	THRLDLPKSRVTCPEVPAEDI.TWODPIRA-GNTDVCREV--VKQIAOSCSISISQVSTA	408
OV		
DD	33/ GAGAGTTPDFFGGGFG--KSPTMLATDLSLRVDFPIYERTIRRWLEHPEELADEFRKAWYKL	411

```
Db 415 IHRDMGPVARYLGPLVPKQTLWQDPPAVSTTSSAKOIASLKSQIRASGLTVSQLSVTA 474
```

QY 495 WUSAKITNGSMMKGGANGAKIRUAPONMOGNEPERLARVLSVIEQLS
 | : | : ||| |||| |||| |||| | : | : | : | : | : | : | : | : | : | :
D6 475 WAAASRRGSDDRGKGANGRIRLOPOVGWEVNDPDGSAAGHSHPFEIOTSEFTRGNIKVS

Q7 553 IADVIVLAGSVGTEKAAKAGYDVRVPELKGGRDATAETMDADSFAPLEPLADGRNWK 612

613 KXXXXXPEMLIJDRAOLMCI TGGPDMTVL I GGMKVI GTNYGGCTHCEVMTDCEGL TINDRTV 672

Db 595 KGNCRPSTSLDKANLLTLSAPETVLGGRLVGLGANYKRLPLGVFTASESLTNDFFV 654

Qy	673	NLTDMGNSWK--	PVGSNAYEIRDRKTCAYKWT	ASRVDLVFGSNLLRSYAEVYAQDDNGE	730
Dh	655	NLTDMGTSFGSAD	NDNNVGVG--	CSGKLVGKNSGDMVTVGCGEVEYDATVGVVA	713

QY 731 KEYRDFVAAWTKYNNADRFV 751

Db 714 KFTGFEAAWDKYMNLDRFDV 734

RESULT 11
AAR08205

AAR08205
 ID AAR08205 standard; protein; 731 AA.

AC AAR08205;

AA
DT 06-MAR-1991 (first entry)

xx DE Heat resistant peroxidase gene product.

XX	pod10; E.coli UM228.
KW	
XX	Bacillus stearothermophilus.
OS	
XX	JP02268684-A.
XX	
PN	02-NOV-1990.
XX	
PD	07-APR-1989; 89JP-0089459.
XX	
PF	07-APR-1989; 89JP-0089459.
XX	
PR	(TOYM) TOYONO KK.
XX	
PA	WPI; 1990-372008/50.
XX	N-PSDB; AAQ06815.
XX	
DR	DNA contg. genetic information of heat resistant peroxidase -
XX	prepd. by culturing transformant and collecting heat resistant
PT	peroxidase
PT	

PS Disclosure; Fig 1; 8pp; Japanese.

CC Heat resistant peroxidase may be produced from a plasmid pod10
CC transformed E.coli UM28 expression system efficiently and
CC in large quantities.

AX	Sequence	731 AA;
----	----------	---------

Query Match	38.1%	Score 1547;	DB 11;	Length 731;
Best Local Similarity	43.9%	Pred. No. 3.4e-125;		
Matches 342; Conservative	89;	Mismatches 228;	Indels 120;	Gaps 18;

Q7 32 RERAMGKCPVMHG--GNTSTGTSTGNDKWPEGINLDILHQQDRKSDPDMDPDENYREEVRK 89

4 QKQKQAAQCFPHESYVINGSSNKTINKDWPNQUNLSILHQHDKRKTNPHEDEEFYAEFFQK 63

Db 64 LDYWALKEDLRKIMTESQDWWZADYGHYGPFLIRMAWHSAGNYRIADGRGGASTGTQRF 123

QY	PLNSWPURVSDUKARRKLLWP	INAKKIGNLSWADLMILAGTVAIESMGLPATG	-----SFGRV	207
				:
				:
D6	PLNSSPPNANLDRKRCYGRSRENTGK	-----SLG-PICSEWRAMSL	166	

QY 208 DIWEPE-----KDIYWGDEKEWLAPSDRYGDNYPETWENPLAAVQMGLI 253
: | : | : | : | : | : | : | : | : |

DD	107	NWWEAKLESAAGFLINSGIKNAIFIGURKKKSGSFLINAI--PVIAIATNKSPRA-----	217
Qv	254	VYNEPEGVNGHPPLRTAOOV-----LETEARMANDEKTAALTAGGHTVYNGCHNGN	305

Db 218 ---NGYNLR-QPRRAGQAGSKSRGISAETFRMGKDEETVALIAGHTFGKAHRGGP 272

QY ASALAPUPKASDVENGLGWSNPNMQGKASNVAVTSGLEGAWTNTPTAFDMGFDDLPGYN 365

Dp AHVHGPRDEAADTEAAOGI GWTSVCKGKGSDMTTSGLEGAWTDTDQWNTSVNDMTSGVN 332

QY 366 WELKSPAGAHHEPIDIKKENKPVDSPIRHNPIINTDADMAIKVNPITYRAICEKEMA 425

DB 333 WVLTKSPAGAWQWVDPDEKDLAPDAEDPSKKVPTMMWTTDLALREPEYER IARRHQ 392

Db 393 NPEEFAEAFARAWFKLTHRDMGPKRYLGPVPKEDFWQDPRIPEVDYELTAEIEEIK 452

QY	483	KIAQGLSISEMYSTANDSARTYRGSDMRGGANGARIRLAPQNEWQNEPERLAKVLSVY	542
QZ	484	WIKWQGLSISEMYSTANDSARTYRGSDMRGGANGARIRLAPQNEWQNEPERLAKVLSVY	543
QZ	485	WIKWQGLSISEMYSTANDSARTYRGSDMRGGANGARIRLAPQNEWQNEPERLAKVLSVY	544

QY 543 EQISADTGASDIADVIVLAGSVGIEKAKAAGYDVRVPFLKG-----RGDATAEMTTDADSF 597

Db 512 RGPARTA-----EKSKHR-----RLDRLGGTLRWKRPATPALMSNCHF 551

XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: the sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX CC Sequence 273 AA;

Query Match 19.3%; Score 784.5; DB 22; Length 273;
Best Local Similarity 58.4%; Pred. No. 1.4e-59;
Matches 160; Conservative 36; Mismatches 73; Indels 5; Gaps 3;

QY 481 KOKIAQSGLSISEMSTAWDSARTYRSGDMRGANGARIRLAPONEWQNEPERLAKVLS 540
DB 1 KFAIADSGLSISEMSTAWDSARTYRSGDMRGANGARIRLAPONEWQNEPERLAKVLS 58
QY 541 VYEQISADTG-ASIAADVIVLAGSVGIEKAAAGYDVRVPFLKRGDGTAEATDADSFAP 599
DB 59 VLEKIOKESGKASLADIIVLAGVGVGKAAAGLSIHVPFAPGRVDAQDQTDIEMFEL 118
QY 600 LEPLADGFRNWKQKVEYVYKPEEMLLDRAQLMGLTGPENTVLLGGHRLVLTGNTGKKGVF 659
DB 119 LEPIADGFRNRYARLDVSTTESLLIDKAOQLTAPENTALVGGMRVLGANFDGKNGVF 178
QY 660 TCEGQLTNDFFVNLTDGMSKPV--GSNAYEIRDRKTGAVKWTASRDVLVFGNSLLR 717
DB 179 TORVGVLNDFVNLDMRYENKATDESKELFEGRDRETGEMKFTASRADLVFGNSVLR 238
QY 718 SYAEVYAADNGEKFVRDFVAAWTVMNADREDV 751
DB 239 AVAEVYASSDAHEKFKVDFVAAWVKVMDLREDL 272

RESULT 14
ABG09297
ID ABG09297 standard; Protein; 1550 AA.
XX AC ABG09297;
XX AC ABG09297;
XX DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #9288.
XX DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX DE food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI: 2001-639362/73.
XX DR N-PSDB; AAS73484.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -
XX PS Claim 20; SEQ ID No 39636; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX CC Sequence 1550 AA;

Query Match 12.8%; Score 519; DB 22; Length 1550;
Best Local Similarity 49.8%; Pred. No. 2.6e-35;
Matches 107; Conservative 34; Mismatches 66; Indels 8; Gaps 5;

QY 350 PKFDMGYDILLFGYNWELKSPAGAHWEPIDIKKENPVDASDSIRHNPIMTDADMA 409
DB 529 PTQWSNYFFENLFKYEWVQTRSPAGAIQFEAVD-APELIP-DPFDPSKKRKTMLVTLT 586
QY 410 IKVNPYRAICEKFMADPEYFKTKAFKAMFKLTHRDLGPKSRVIGPEVPAEDLIWQDPI 469
DB 587 LRFDPFEKISRRLNDPQAFNEAFARAFWFKLTHRDMGPKSRVIGPEVPAEDLIWQDPI 646
QY 470 AGNTDYCEEVW---KOKIAQSGLSISEMSTAWDSARTYRSGDMRGANGARIRLAPONE 526
DB 647 QPIYNTEQDIIDLKFAIADSGLSISEMSTAWDSARTYRSGDMRGANGARIRLAPONE 706
QY 527 WQNEPERLAKVLSVYEQISADTG-ASIAADVIVLA 560
DB 707 WDVN--AAAVRALPVLEKIOKESGKASLADIIVLA 739

RESULT 15
ABG08828
ID ABG08828 standard; Protein; 201 AA.
XX AC ABG08828;
XX AC ABG08828;
XX DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #8819.
XX DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX DE food supplement; medical imaging; diagnostic; genetic disorder.
XX

```

OS Homo sapiens.
XX WO200175067-A2.
XX
XX PD
XX 11-OCT-2001.
XX
XX PF 30-MAR-2001; 2001WO-US08631.
XX
XX PR 31-MAR-2000; 2000US-0540217.
XX
XX PR 23-AUG-2000; 2000US-0649167.
XX
XX PA (HYSE-) HYSEQ INC.
XX
XX PI Drmanac RT, Liu C, Tang YT;
XX
XX DR WPI; 2001-639362/73.
XX
XX DR N-PSDB; AAS73015.
XX
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX
XX PS Claim 20; SEQ ID No 39187; 103pp; English.
XX
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABC30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 201 AA;

Query Match 9.88; Score 398; DB 22; Length 201;
Best Local Similarity 41.2%; Pred. No. 3.6e-26;
Matches 89; Conservative 24; Mismatches 43; Indels 60; Gaps 3;

QY 506 RGSMDRGANGARLAPQ---NEWOGNEPERLAKVLSVEQISADTGASIAADVIVLAGS 562
Db 21 RGGDKRGANGARLAPQRTGDTVRQPIRVCTVIEMEL----- 62

QY 563 VGIEKAAAGYDVRVFLKRGDATAEMTDADSFAPLEPLADGFRNWKKEYVVKPEEM 622
Db 63 -----LEPIADGFRNVRARLDVSTTESL 85

QY 623 LLDRAQLMGLTGPMVTLLGMRVLTGNYGTHGVFTDCGQLTNDFFVNLDMGNSWK 682
Db 86 LIDRAQQLTTEAPETALVGGMRYLGANFDGSKNGVFTDRVGVLSNDDFFVNLDMRYEWK 145

QY 683 PV--GSNAYEIRDRKGTGAVKWTASRVDLVFGNSLL 716
Db 146 ATDESKELFEGRDRETGEVFTASRADLVFGNSVL 181

```

Search completed: October 7, 2003, 19:42:20
Job time : 64.8231 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 7, 2003, 19:40:18 ; Search time 22.1175 Seconds
(without alignments)
1440.492 Million cell updates/sec

Title: US-09-884-889-6
Perfect score: 4058
Sequence: 1 MNASADLHSSLQRCRAF.....RDFVAANTKVNADRFVAS 753

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: , 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4058	100.0	753	2 US-08-674-887A-6	Sequence 6, Appl
2	4058	100.0	753	2 US-08-951-844-6	Sequence 6, Appl
3	4058	100.0	753	4 US-09-412-347-6	Sequence 6, Appl
4	2290.5	56.4	745	2 US-08-674-887A-8	Sequence 8, Appl
5	2290.5	56.4	745	3 US-08-951-844-8	Sequence 8, Appl
6	2290.5	56.4	745	4 US-09-412-347-8	Sequence 8, Appl
7	2093	51.6	726	2 US-08-313-185-49	Sequence 49, Appl
8	2093	51.6	726	2 US-08-459-499-13	Sequence 13, Appl
9	2093	51.6	726	3 US-09-082-614A-49	Sequence 49, Appl
10	2072.5	51.1	726	4 US-09-328-352-8115	Sequence 8115, Ap
11	2044.5	50.4	740	1 US-08-418-782-21	Sequence 21, Appl
12	2044.5	50.4	740	2 US-08-852-219-21	Sequence 21, Appl
13	2041.5	50.3	740	1 US-08-418-782-7	Sequence 7, Appl
14	2041.5	50.3	740	1 US-08-228-662-7	Sequence 7, Appl
15	2041.5	50.3	740	2 US-08-852-219-7	Sequence 7, Appl
16	2040.5	50.3	729	2 US-08-313-185-50	Sequence 50, Appl
17	2040.5	50.3	729	2 US-08-459-499-14	Sequence 14, Appl
18	2040.5	50.3	729	3 US-09-082-614A-50	Sequence 50, Appl
19	1890	46.6	735	2 US-08-313-185-48	Sequence 48, Appl
20	1890	46.6	735	2 US-08-459-499-9	Sequence 9, Appl
21	1890	46.6	735	3 US-09-082-614A-48	Sequence 48, Appl
22	1885	46.5	735	2 US-08-459-499-12	Sequence 12, Appl
23	1681	41.4	652	2 US-08-313-185-53	Sequence 53, Appl
24	1681	41.4	652	2 US-08-459-499-17	Sequence 17, Appl
25	1681	41.4	652	3 US-09-082-614A-53	Sequence 53, Appl
26	1616	39.8	731	2 US-08-313-185-51	Sequence 51, Appl
27	1616	39.8	731	2 US-08-459-499-15	Sequence 15, Appl

28 1616 39.8 731 3 US-09-082-614A-51 Sequence 51, Appl
29 243.5 6.0 294 2 US-08-459-499-16 Sequence 16, Appl
30 242 6.0 293 2 US-08-313-185-52 Sequence 52, Appl
31 242 6.0 293 3 US-09-082-614A-52 Sequence 52, Appl
32 188 4.6 37 1 US-07-929-206-6 Sequence 6, Appl
33 188 4.6 37 2 US-08-313-185-3 Sequence 3, Appl
34 188 4.6 37 2 US-08-459-499-6 Sequence 6, Appl
35 188 4.6 37 3 US-09-082-614A-3 Sequence 3, Appl
36 180.5 4.4 78 1 US-07-929-206-4 Sequence 4, Appl
37 180.5 4.4 78 2 US-08-313-185-44 Sequence 44, Appl
38 180.5 4.4 78 2 US-08-459-499-4 Sequence 4, Appl
39 180.5 4.4 78 3 US-09-082-614A-44 Sequence 44, Appl
40 180 4.4 37 1 US-07-929-206-5 Sequence 5, Appl
41 180 4.4 37 1 US-07-929-206-8 Sequence 8, Appl
42 180 4.4 37 2 US-08-313-185-2 Sequence 2, Appl
43 180 4.4 37 2 US-08-459-499-5 Sequence 5, Appl
44 180 4.4 37 3 US-09-082-614A-2 Sequence 2, Appl
45 156 3.8 120 1 US-07-929-206-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-674-887A-6
; Sequence 6, Application US/08674887A
; Patent No. 5939300
; GENERAL INFORMATION:
; APPLICANT: Robertson, Dan E.
; APPLICANT: Sanyal, Indrajit
; APPLICANT: Adhikari, Robert S.
; TITLE OF INVENTION: CATALASES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/674,887A
; FILING DATE: 03-JUL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09015/002001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 753 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-674-887A-6

Query Match 100.0%; Score 4058; DB 2; Length 753;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 753; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNASADLHSSLQRCRAFVPLVSPRRAIRRAMSGKCPVMHGNTSTGTGNKDWPE 60
|||||

```

Db 1 MNNASADDLHSSLQORCRAFPVLPSPRHAIIRERAMSGKCPVMHGGNTSTGTSNKNWPE 60
Qy 61 GLNLDIHQDQKSDPMDPDFNYREYEVKRLDFDALKDVHALMTDSQEWNPADWGHYGL 120
Db 61 GLNLDIHQDQKSDPMDPDFNYREYEVKRLDFDALKDVHALMTDSQEWNPADWGHYGL 120
Qy 121 MIRMAHSAGTYRIADRGGGGTGSRFAPLNSPNDVSLDKARLLWPIKKNYGNKISW 180
Db 121 MIRMAHSAGTYRIADRGGGGTGSRFAPLNSPNDVSLDKARLLWPIKKNYGNKISW 180
Qy 181 ADLMILAGTVAYESMGLPAYGFSRGRVDIWEPEKDIYWGDEKELAPSDERYGDNKPT 240
Db 181 ADLMILAGTVAYESMGLPAYGFSRGRVDIWEPEKDIYWGDEKELAPSDERYGDNKPT 240
Qy 241 MENPLAAVQMLIIYVNPPEGVNGHPDPLRTAQOVLTFARMANDEKTAALTAGHTVNC 300
Db 241 MENPLAAVQMLIIYVNPPEGVNGHPDPLRTAQOVLTFARMANDEKTAALTAGHTVNC 300
Qy 301 HGNGNASALAPDPKASDVENOGLGWGNPNMOGKASNAVTSIGEGAWTNTPTKFDMGYFDL 360
Db 301 HGNGNASALAPDPKASDVENOGLGWGNPNMOGKASNAVTSIGEGAWTNTPTKFDMGYFDL 360
Qy 361 LFQYNWELKSPAGAHWEPIIDIKENKPVDSAPDSIRHNPIMTDADMAIKVNPYRAIC 420
Db 361 LFQYNWELKSPAGAHWEPIIDIKENKPVDSAPDSIRHNPIMTDADMAIKVNPYRAIC 420
Qy 421 EKFMADPEYFKKTFKAKWFKLTHRDGLPKSRYIGPEVPAEDLIWQDPIPAAGTDCYEV 480
Db 421 EKFMADPEYFKKTFKAKWFKLTHRDGLPKSRYIGPEVPAEDLIWQDPIPAAGTDCYEV 480
Qy 481 KQKIAQSGLSISEMVSSTANDSARTYRGSDMRGGANGARIRLAPONEQWGPPELAKVLS 540
Db 481 KQKIAQSGLSISEMVSSTANDSARTYRGSDMRGGANGARIRLAPONEQWGPPELAKVLS 540
Qy 541 VYEQISADTGASITADVIIVLAGSVGIEKAAKAAGYDVRVPLKGRGDATAEMTDADSPAPL 600
Db 541 VYEQISADTGASITADVIIVLAGSVGIEKAAKAAGYDVRVPLKGRGDATAEMTDADSPAPL 600
Qy 601 EPLADGFRNQKKEYYVVKPEMLLDRAQLMGLTGPEMTVLGGMRVLGTNYGKTHGVT 660
Db 601 EPLADGFRNQKKEYYVVKPEMLLDRAQLMGLTGPEMTVLGGMRVLGTNYGKTHGVT 660
Qy 661 DCEGQLTNDFVNLTDMGNSWKPVGSNAYEIRDRKTKGAVKWTASRVDLVFGSNLLRSYA 720
Db 661 DCEGQLTNDFVNLTDMGNSWKPVGSNAYEIRDRKTKGAVKWTASRVDLVFGSNLLRSYA 720

```

```

RESULT 2
US-08-951-844-6
; Sequence 6, Application US/08951844
; Patent No. 6074860
; GENERAL INFORMATION:
; APPLICANT: Robertson et al.
; TITLE OF INVENTION: Catalogs
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/951.844
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/674,887
; FILING DATE: July 3, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles J. Herron
; REGISTRATION NUMBER: 28,019
; REFERENCE/DOCKET NUMBER: 331400-55
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 753 AMINO ACIDS
; TYPE: AMINO ACID
; TOPOLOGY: LINEAR
; MOLECULE TYPE: Polypeptide
; US-08-951-844-6

```

```

Query Match 100.0%; Score 4058; DB 3; Length 753;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 753; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MNNASADDLHSSLQORCRAFPVLPSPRHAIIRERAMSGKCPVMHGGNTSTGTSNKNWPE 60
Db 1 MNNASADDLHSSLQORCRAFPVLPSPRHAIIRERAMSGKCPVMHGGNTSTGTSNKNWPE 60
Qy 61 GLNLDIHQDQKSDPMDPDFNYREYEVKRLDFDALKDVHALMTDSQEWNPADWGHYGL 120
Db 61 GLNLDIHQDQKSDPMDPDFNYREYEVKRLDFDALKDVHALMTDSQEWNPADWGHYGL 120
Qy 121 MIRMAHSAGTYRIADRGGGGTGSRFAPLNSPNDVSLDKARLLWPIKKNYGNKISW 180
Db 121 MIRMAHSAGTYRIADRGGGGTGSRFAPLNSPNDVSLDKARLLWPIKKNYGNKISW 180
Qy 181 ADLMILAGTVAYESMGLPAYGFSRGRVDIWEPEKDIYWGDEKELAPSDERYGDNKPT 240
Db 181 ADLMILAGTVAYESMGLPAYGFSRGRVDIWEPEKDIYWGDEKELAPSDERYGDNKPT 240
Qy 241 MENPLAAVQMLIIYVNPPEGVNGHPDPLRTAQOVLTFARMANDEKTAALTAGHTVNC 300
Db 241 MENPLAAVQMLIIYVNPPEGVNGHPDPLRTAQOVLTFARMANDEKTAALTAGHTVNC 300
Qy 301 HGNGNASALAPDPKASDVENOGLGWGNPNMOGKASNAVTSIGEGAWTNTPTKFDMGYFDL 360
Db 301 HGNGNASALAPDPKASDVENOGLGWGNPNMOGKASNAVTSIGEGAWTNTPTKFDMGYFDL 360
Qy 361 LFQYNWELKSPAGAHWEPIIDIKENKPVDSAPDSIRHNPIMTDADMAIKVNPYRAIC 420
Db 361 LFQYNWELKSPAGAHWEPIIDIKENKPVDSAPDSIRHNPIMTDADMAIKVNPYRAIC 420
Qy 421 EKFMADPEYFKKTFKAKWFKLTHRDGLPKSRYIGPEVPAEDLIWQDPIPAAGTDCYEV 480
Db 421 EKFMADPEYFKKTFKAKWFKLTHRDGLPKSRYIGPEVPAEDLIWQDPIPAAGTDCYEV 480
Qy 481 KQKIAQSGLSISEMVSSTANDSARTYRGSDMRGGANGARIRLAPONEQWGPPELAKVLS 540
Db 481 KQKIAQSGLSISEMVSSTANDSARTYRGSDMRGGANGARIRLAPONEQWGPPELAKVLS 540
Qy 541 VYEQISADTGASITADVIIVLAGSVGIEKAAKAAGYDVRVPLKGRGDATAEMTDADSPAPL 600
Db 541 VYEQISADTGASITADVIIVLAGSVGIEKAAKAAGYDVRVPLKGRGDATAEMTDADSPAPL 600
Qy 601 EPLADGFRNQKKEYYVVKPEMLLDRAQLMGLTGPEMTVLGGMRVLGTNYGKTHGVT 660
Db 601 EPLADGFRNQKKEYYVVKPEMLLDRAQLMGLTGPEMTVLGGMRVLGTNYGKTHGVT 660
Qy 661 DCEGQLTNDFVNLTDMGNSWKPVGSNAYEIRDRKTKGAVKWTASRVDLVFGSNLLRSYA 720
Db 661 DCEGQLTNDFVNLTDMGNSWKPVGSNAYEIRDRKTKGAVKWTASRVDLVFGSNLLRSYA 720

```

QY 721 EYVADDDNGEKEFVRDFAAATKVMNADREFDVAS 753
Db 721 EYVADDDNGEKEFVRDFAAATKVMNADREFDVAS 753

RESULT 3

US-09-412-347-6
; Sequence 6, Application US/09412347
; Patent No. 6410290
; GENERAL INFORMATION:
; APPLICANT: Robertson, Dan E.
; APPLICANT: Sanyal, Indrajit
; APPLICANT: Adhikari, Robert S.
; TITLE OF INVENTION: CATALASES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/412,347
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/674,887
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Halle, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09015/002001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 753 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-09-412-347-6

Query Match 100.0%; Score 4058; DB 4; Length 753;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 753; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNASADLLHSSLOQCRAFPVLVSPRRAIRERAMSGKCPVHGGTSTGTSGNKDWPE 60
Db 1 MNASADLLHSSLOQCRAFPVLVSPRRAIRERAMSGKCPVHGGTSTGTSGNKDWPE 60
QY 61 GLNLDLHQDRKSDPMDPFNYREVRKLDLFDALKDVKVHALMTDSQEWNPADWGHYGG 120
Db 61 GLNLDLHQDRKSDPMDPFNYREVRKLDLFDALKDVKVHALMTDSQEWNPADWGHYGG 120
QY 121 MIRMAHWSAGTVRIADGRGGGTGSGRFPAPLNSWPNVSLDKARRLLWPIKKYGNKISW 180
Db 121 MIRMAHWSAGTVRIADGRGGGTGSGRFPAPLNSWPNVSLDKARRLLWPIKKYGNKISW 180
QY 181 ADMILLAGTVAVESMGLPAYGFSFGVDIWEPEKDIYWGDEKWLAPSDRYGVNKPET 240
Db 181 ADMILLAGTVAVESMGLPAYGFSFGVDIWEPEKDIYWGDEKWLAPSDRYGVNKPET 240
QY 241 MENPLAAVQMGILYVNPVGVNHPDPLRTAQOVLFTFARMAMNDEKTAALTAGGHTVGN 300
Db 241 MENPLAAVQMGILYVNPVGVNHPDPLRTAQOVLFTFARMAMNDEKTAALTAGGHTVGN 300

QY 301 HGNGNASALAPDPKASDVENQGLGNGNPNMOGKASNAVTSIGIEGAWTNTPTKFDMGYFDL 360
Db 301 HGNGNASALAPDPKASDVENQGLGNGNPNMOGKASNAVTSIGIEGAWTNTPTKFDMGYFDL 360
QY 361 LFGYNWELKKSAPAGAHWEPIIDIKKENKPVDA SDPSIRHNPIWTDADMAIKVNPYRAIC 420
Db 361 LFGYNWELKKSAPAGAHWEPIIDIKKENKPVDA SDPSIRHNPIWTDADMAIKVNPYRAIC 420
QY 421 EKFMADPEYFKTKTFAKAWFKLTHRD LGPKSRYIGPEVPAEDLIWODPI PAGNTDCEEVY 480
Db 421 EKFMADPEYFKTKTFAKAWFKLTHRD LGPKSRYIGPEVPAEDLIWODPI PAGNTDCEEVY 480
QY 481 KOKIAQSGLSISEMSTAWSDARTYRGSDMRGANGARIRLAPQNEWQNEPERLAKVLS 540
Db 481 KOKIAQSGLSISEMSTAWSDARTYRGSDMRGANGARIRLAPQNEWQNEPERLAKVLS 540
QY 541 VYEQISADTGASIDVIVLAGSVGTEKAAAGYDVRVPEFLKRGDATAEMTDADSFAPL 600
Db 541 VYEQISADTGASIDVIVLAGSVGTEKAAAGYDVRVPEFLKRGDATAEMTDADSFAPL 600
QY 601 EPLADGFRNWQKKEYVVKPEEMLLDRAQLMGLTGPEMTVLLGGMRVLGTYNGGTRHGVFT 660
Db 601 EPLADGFRNWQKKEYVVKPEEMLLDRAQLMGLTGPEMTVLLGGMRVLGTYNGGTRHGVFT 660
QY 661 DCEGQLTNDFFVNLTDGMSWKPVGSNAYEIRDRKTGAVKWTASRVDLVFGSNSLLRSYA 720
Db 661 DCEGQLTNDFFVNLTDGMSWKPVGSNAYEIRDRKTGAVKWTASRVDLVFGSNSLLRSYA 720
QY 721 EYVADDDNGEKEFVRDFAAATKVMNADREFDVAS 753
Db 721 EYVADDDNGEKEFVRDFAAATKVMNADREFDVAS 753

RESULT 4

US-08-674-887A-8
; Sequence 8, Application US/08674887A
; Patent No. 5939300
; GENERAL INFORMATION:
; APPLICANT: Robertson, Dan E.
; APPLICANT: Sanyal, Indrajit
; APPLICANT: Adhikari, Robert S.
; TITLE OF INVENTION: CATALASES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/674,887A
; FILING DATE: 03-JUL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Halle, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09015/002001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 745 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
us-08-674-887A-8

Query Match 56.4%; Score 2290.5; DB 2; Length 745;
Best Local Similarity 59.1%; Pred. No. 1.8e-207;
Matches 433; Conservative 92; Mismatches 183; Indels 25; Gaps 8;

Qy 38 GKCPVHGG---NTSTGTSNKKDMPGEGILDLHQDQKSDPMDPFDNYREEVKRLDFA 94
Db 18 GKCPFTGSLKQSGAGGKKNRDMWPNMNLGILRHSSLSDPNDPFDYAEFFKDLAA 77

Qy 95 LKDVHMLTDSQWNPADMGHYGGLMIRMAHWSAGTYRIADGRGGGTSQRFAPLNSW 154
Db 78 VKKDLAALMTDSQWNPADMGHYGGLMIRMAHWSAGTYRIADGRGGGTSQRFAPLNSW 137

Qy 155 PDNVSLDKARLLWPIKKGKISWADLMTAGTVAYESMGLPAYGFSRGRVDIWPPEK 214
Db 138 PDNANLDKARLLWPIKKGKISWADLMTAGTVAYESMGLPAYGFSRGRVDIWPPEK 197

Qy 215 DIYWGDEKWLAPSDERY-GDVNKPETMENPLAAVQMGILYVNPPEGVNGHPDPLRTAQV 273
Db 198 DVYWGAEWML-GDKRYEGD-----RELENPLGAVQMGILYVNPPEGVNGHPDPLRTAQV 251

Qy 274 LETFARMAMDEKTAALTAGHTVGNCHGNASALAPDPKASDVNQGGLWGNPNMQG 332
Db 252 RETFARMAMDEETVALLAGHTFGKTHGAADAEKYVGREPAAAGIEMSLGWKNTYGTG 311

Qy 333 KASNAVTSIGEGAWTTPNPTKFDGMYFDLLFGYNELKSPAGAUHWSPIDIKKENKPVDA 392
Db 312 HGADTITSGLEGAWTKPTQWSNFFENLFGYEWELTKSPAGAYQWPKDGAGAGTIPDA 371

Qy 393 SDPSIRNHPIMTDADMAIKVNPYTRAIKKEFMADPEYFKTKFAKAWFKLTHRDIGPKSRY 452
Db 372 HDPSKSHAPFMTTDLALRMDPDYKISRYYENPDEFADAFKAWKYLTHRDGPKVRY 431

Qy 453 IGPEVPAEDLIWQDPIAGNTDYCE----EVRKQIAQSGLSISEMSTANDSARTYRGS 508
Db 432 LGPEVPOEDLIWQDPIAGNTDYCE----EVRKQIAQSGLSISEMSTANDSARTYRGS 491

Qy 509 DMRGANGARIRLAPQNEWQNEPERLAKVLSVYEQISAD-----TGASIAADVILA 560
Db 492 DKRGANGARIRLAPQNEWQNEPERLAKVLSVYEQISAD-----TGASIAADVILA 551

Qy 561 GSVGIEKAAKAAGYDVRVPLKGRGDATAEMTDADSPAPLEPLADGFRNNOKKEYVYKPE 620
Db 552 GCAGVEKAAKADAGHEVQVPPNPGRADATASQTDVEAFEALEPAADGFRNTPKPKYSAE 611

Qy 621 EMLLDRAQLMGLTGPEMTVLLGGHVRVLTNYGKTHGVFTDCGQLTNDFFVNLTMGNS 680
Db 612 EMLVDRAQLLSAPEMTALVGGHVRVLTNYGKTHGVFTDCGQLTNDFFVNLTMGNS 671

Qy 681 WK--PVGSNAYEIRDRKTGAKWTASRVLDVFGSNLSLRSAEYVAQDDNGEKVRDFVA 738
Db 672 WRASDESCKVFEGRDFTGKSGTRVLDLFGSNLSLRSAEYVAQDDNGEKVRDFVA 731

Qy 739 ATKVMNADRPDV 751
Db 732 AWAKVMDLRDFDL 744

RESULT 5

US-08-951-844-8
; Sequence 8, Application us/08951844
; Patent No. 6074860
; GENERAL INFORMATION:
; APPLICANT: Robertson et al.
; TITLE OF INVENTION: Catalogs
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD

; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/951,844
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/674,887
; FILING DATE: July 3, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles J. Heiron
; REGISTRATION NUMBER: 28,019
; REFERENCE/DOCKET NUMBER: 331400-55
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 745 AMINO ACIDS
; TYPE: AMINO ACID
; TOPOLOGY: LINEAR
; MOLECULE TYPE:
US-08-951-844-8

Query Match 56.4%; Score 2290.5; DB 3; Length 745;
Best Local Similarity 59.1%; Pred. No. 1.8e-207;
Matches 433; Conservative 92; Mismatches 183; Indels 25; Gaps 8;

Qy 38 GKCPVHGG---NTSTGTSNKKDMPGEGILDLHQDQKSDPMDPFDNYREEVKRLDFA 94
Db 18 GKCPFTGSLKQSGAGGKKNRDMWPNMNLGILRHSSLSDPNDPFDYAEFFKDLAA 77

Qy 95 LKDVHMLTDSQWNPADMGHYGGLMIRMAHWSAGTYRIADGRGGGTSQRFAPLNSW 154
Db 78 VKKDLAALMTDSQWNPADMGHYGGLMIRMAHWSAGTYRIADGRGGGTSQRFAPLNSW 137

Qy 155 PDNVSLDKARLLWPIKKGKISWADLMTAGTVAYESMGLPAYGFSRGRVDIWPPEK 214
Db 138 PDNANLDKARLLWPIKKGKISWADLMTAGTVAYESMGLPAYGFSRGRVDIWPPEK 197

Qy 215 DIYWGDEKWLAPSDERY-GDVNKPETMENPLAAVQMGILYVNPPEGVNGHPDPLRTAQV 273
Db 198 DVYWGAEWML-GDKRYEGD-----RELENPLGAVQMGILYVNPPEGVNGHPDPLRTAQV 251

Qy 274 LETFARMAMDEKTAALTAGHTVGNCHGNASALAPDPKASDVNQGGLWGNPNMQG 332
Db 252 RETFARMAMDEETVALLAGHTFGKTHGAADAEKYVGREPAAAGIEMSLGWKNTYGTG 311

Qy 333 KASNAVTSIGEGAWTTPNPTKFDGMYFDLLFGYNELKSPAGAUHWSPIDIKKENKPVDA 392
Db 312 HGADTITSGLEGAWTKPTQWSNFFENLFGYEWELTKSPAGAYQWPKDGAGAGTIPDA 371

Qy 393 SDPSIRNHPIMTDADMAIKVNPYTRAIKKEFMADPEYFKTKFAKAWFKLTHRDIGPKSRY 452
Db 372 HDPSKSHAPFMTTDLALRMDPDYKISRYYENPDEFADAFKAWKYLTHRDGPKVRY 431

Qy 453 IGPEVPAEDLIWQDPIAGNTDYCE----EVRKQIAQSGLSISEMSTANDSARTYRGS 508
Db 432 LGPEVPOEDLIWQDPIAGNTDYCE----EVRKQIAQSGLSISEMSTANDSARTYRGS 491

Qy 509 DMRGANGARIRLAPQNEWQNEPERLAKVLSVYEQISAD-----TGASIAADVILA 560
Db 492 DKRGANGARIRLAPQNEWQNEPERLAKVLSVYEQISAD-----TGASIAADVILA 551

Qy 561 GSVGIEKAAKAAGYDVRVPLKGRGDATAEMTDADSPAPLEPLADGFRNNOKKEYVYKPE 620
Db 552 GCAGVEKAAKADAGHEVQVPPNPGRADATASQTDVEAFEALEPAADGFRNTPKPKYSAE 611

Qy 621 EMLLDRAQLMGLTGPEMTVLLGGHVRVLTNYGKTHGVFTDCGQLTNDFFVNLTMGNS 680
Db 612 EMLVDRAQLLSAPEMTALVGGHVRVLTNYGKTHGVFTDCGQLTNDFFVNLTMGNS 671

Qy 681 WK--PVGSNAYEIRDRKTGAKWTASRVLDVFGSNLSLRSAEYVAQDDNGEKVRDFVA 738
Db 672 WRASDESCKVFEGRDFTGKSGTRVLDLFGSNLSLRSAEYVAQDDNGEKVRDFVA 731

Qy 739 ATKVMNADRPDV 751
Db 732 AWAKVMDLRDFDL 744

552	CCACGYEKAADGHEGVQFPNPGRADATAEQTVDEAFEALEPAADGFRNVIPEHKVSAE	611
621	EMLLDRAQLMGLTGPEMTVLILGGMRLVLTGYGTTKGHVFTDCBGQLTDFVNLTDGMS	680
612	EMLDVRAQLLSAPENTALVGGMRVLGTNYDGSQHGVTNKGQSLNDFVNLDDLNTK	671
681	WK--PVGSNAYEIRDRKTGAYKWTASRVLDVFGSNLSLRSYAEYAQDDNGSGEVRDFA	738
672	WRASDEKVFEGRDFKTGTEGKWSGTRVDLLFGSNSELRALAEVYGCADSEEEKFYKDFVK	731
739	AWTKVNNADREDV	751
732	AAAKVMDLDRFDL	744

RESULT 6

```

US-09-412-347-8
; Sequence 8, Application US/09412347
; Patent No. 6410290
; GENERAL INFORMATION:
; APPLICANT: Robertson, Dan E.
; APPLICANT: Sanyal, Indrajit
; APPLICANT: Adhikari, Robert S.
; TITLE OF INVENTION: CATALASES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: Fastseq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/412,347
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/674,887

```

Query Match	56.4%	Score 2290.5	DB 4	Length 745
Best Local Similarity	59.1%	Pred. No. 1.8e-207		
Matches 433	Conservative 92	Mismatches 183	Indels 25	Gaps 8
Qy	38	GKCPVMHGG---NTSTGTSNKDWPEGLNLDLILHQDQRKSDPMDPDFNYREEVKKLDFDA	94	
Db	18	GKCPFTGGSLKQSAGCGTKNRDWWPNMLNLGILRQHSSLSDPNDPDFDAEEFKKLDA	77	
Qy	95	LKDKVTALWTDSEWNPADMGHYGGLMIRMAWHSAGTYRIADRGCGGTCGSRFAPLNSW	154	
Db	78	VKKDLAALWTDSDQWNPADYGHYGPFFIRMAWHSAGTYRIGDRGGGSGCSGRFAPLNSW	137	
Qy	155	PDNVSLDKARLLWPTKKTKYGNKISWADLMILACTVAYESMGUYPAYGFSGRVDIWPPEK	214	

138	Db	PDNANDKARLLWPIKQYGRKISWADLMILTGNVALETWGMFKTFGFAGGRADWPEE	197
215	QY	DIYWGDEKEWAPSDERY - GDVKNPWTENPAAVQMGLIYVNPVGNGHPDLRTAQOV	273
198	Db	LEWGAETEWL - GDKRYEGD - - - RELENPGLAVQMGLIYVNPVGNGKPDPIAARDI	251
274	QY	LETFARMANDEKTAALTAGCHTVGNGCHNGNASH - LAPDPKASDVENOGLGWGNPNMOG	332
252	Db	RETFFGRMANDEETVALIAGHTFGKTHGAADAKEYVGREPAAAGIECMSLGSWKNTYGTG	311
333	QY	KASNAVTSIGTEGAWTTNPTKDMGYSFDLLFGYNNELKKSAPAGAHUWEPIDIKKENKPYDA	392
312	Db	HGADTTSIGLEGAWTKTPTQWNNFFENLFGYEWELTKSPAGAYQMKPDGAGATIPDA	371
393	QY	SDPSIRHNPLMTDADMAIKVNTPTYRAICEKFMADPEYFKKTFKAWFKLTHRLDGPKSRY	452
372	Db	HDPSKSHAPMLTTDLALRMDPDYEKISRYYENPDEFADAFKAWYKLTHRDMGPKVRY	431
453	QY	IGPEVPAEDLIWQDPIPAAGNTDYCE - - - EYVKOKIAQSGLSISEMVSATMSDARTYRGS	508
432	Db	LGPEVPOEDLIWQDPIPDVSHPLVDENDIEGLKAKILESGLTSELVSTAWASATFRNS	491
509	QY	DMRGANGARIRLAPQEWQGNQNEPERLAKVLVSYEQISAD - - - - - TGASIAADVILA	560
492	Db	DKRGGANGARIRLAPQDNWYNNPQQLARVLKLTLEGIOEDNQAO SDKNAVSLADILIVA	551
561	QY	GSYGIEKAAKAAGYDVRVPFLKRGDATAEMTDADSFAPLEPLADGFRNWKKEYVWKPE	620
552	Db	GCAGVEKAAKDAGHEVQVFPNPGRADATAEQTDVEAFEALPEADGFRNYIKPEHKVSAE	611
621	QY	EMLLDRAQLMGLTGPMTYVLLGGMRVLTNTYGGTKHGVFTQCEGQLTWNDFVNLTDGNS	680
612	Db	EMLVDRAOQLSLSPAPENTALYGGMRVLTNTYDGSQHGVFTKPKGQLSNDFVNLJDLNLT	671
681	QY	WK - - - PVGSNAYEIRDRKTGAVKWTASRVLDLVFGSNLLRSYAETVAQDDNCEKPYRDFVA	738
672	Db	WRASDESQKVFEGRDFDTGVEKVKWGSRTVDLITFGSNSELRALAEVTCADSEKFKVDFVK	731
739	QY	AWTKVMNADRFDV	751
732	Db	AWAKVMDLDRFDL	744

RESULT 7

```

US-08-313-185-49
; Sequence 49, Application US/08313185
; Patent No. 5851763
; GENERAL INFORMATION:
; APPLICANT: Heym, Beate
; APPLICANT: Cole, Stewart
; APPLICANT: Young, Douglas
; APPLICANT: Zhang, Ying
; APPLICANT: Honore, Nadine
; APPLICANT: Telenti, Amalio
; APPLICANT: Bodmer, Thomas
; TITLE OF INVENTION: Rapid Detection of Antibiotic Resistance
; TITLE OF INVENTION: in Mycobacterium Tuberculosis
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313.185

```

```

; FILING DATE: 12-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 02356.0068-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 726 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-313-185-49

```

```

Query Match 51.6%; Score 2093; DB 2; Length 726;
Best Local Similarity 53.4%; Pred. No. 8e-189;
Matches 405; Conservative 103; Mismatches 206; Indels 44; Gaps 10;

```

```

QY 4 ASADLHSSLQRCRAFPVLPSPHRAIRERAMSGKCPVMHGG---NTSTGTSNKDWPE 60
DB 2 STSDDIHT-----TATGKCPFHGGHDQSGAGATTTTRDMPN 39
QY 61 GLNLDTLHQDRKSDPMDPDFNYREEVRKLDLFDALKKDVHMLTDSQEWMPADMGHYGGL 120
DB 40 QLRVDDLNOHNSRNPGLGEDFYRKEFSKLDYGLKDLKALLETESQWPPADMGSGYAGL 99
QY 121 MIRMAHSGAGTYRIADRGGGGTGSGORFAPLNSWPDVSLDKARRLLWPDKKYGKISW 180
DB 100 FIRMAHSGAGTYRISIDRGGAGRGQRFAPLNSWPDVSLDKARRLLWPDKKYGKISW 159
QY 181 ADLMILAGTVAYESMGLPAYGFSRGRVDIWEPEKDIYWGDEKELASDERYGVNKPET 240
DB 160 ADLEILAGTVAYESMGLPAYGFSRGRVDIWEPEKDIYWGDEKELASDERYGVNKPET 211
QY 241 M-ENPLAIVQGLIYVPEGVNGHPDPLRTAQVLETFARMAMDEKTAALTAGHTVGN 299
DB 212 LAKAPLGATEMGLIYVPEGVNGHPDPLRTAQVLETFARMAMDEKTAALTAGHTVGN 271
QY 300 CHNGNLSALAPDKASDVNNOGLGWNQKASNAVTSIGTEGATTTPTKFDMGYFD 359
DB 272 THGAGPSTNVPDPEAAPIEQGLGWASTYSGYGVADAITSGLEVWVTQPTOWSNYFE 331
QY 360 LLFCYNWELKSPAGAHWPIDILKENKPVDSADPSIRHNPIMTDADMAIKVNPYRAI 419
DB 332 NLFEYEWQVSPAGATQFAVD-APEIIP-DFPDPSKKRPTMLVTDLTLRFDPEPEKI 389
QY 420 CEKFMADPEYFKKTFKAWPKLTHRLGPKSRYSYIGPEVPAEDLIWQDPIPAQNTDYCEE 479
DB 390 SRRFLNDPQAFNEAFARAWFKLTHRDGPKSRYSYIGPEVPAEDLIWQDPIPAQNTDYCEE 449
QY 480 V---KQIASGLSISWSTANDSARTYSGMRGGANGARIRLAPONEWQGNPEPRLA 536
DB 450 IIDLKFAIDAGSLVSLVAVASASTFRGGKRGANGARLALMPQRDWDVN--AAAV 507
QY 537 KVLSSVYBOISADTG-ASIAIDVILVAGSVGTEKAAKAGYDVRYPFLKRGDGAETMDAD 595
DB 508 RALPVLKIQESKGLADIIVLAVGVVGVKKAASAGLSIHVPFAPGRVDARQDQDIE 567
QY 596 SPAPLEPLADGFRNRQKKEYVVKPEEMLLDRAQLMGLTGPETVLLGMRVLGTNYGGTK 655
DB 568 MFELLEPLADGFRNRLDVSTTESLLDKAQQTLTAPETALVGMVLSGVPDGSK 627
QY 656 HGVTDCGQLTNDFFVNLDMGNSKVPV--GSAVEIRDKKYGAVKWTASRDVLVFGSN 713
DB 628 NGVFTDRVGLSNDFFVNLDMGNSKVPV--GSAVEIRDKKYGAVKWTASRDVLVFGSN 687
QY 714 SLRSYVAEYVADNDNGEKVFRDVFVAAWTKVMNADRFDV 751
DB 688 SVLRVAEYVADNDNGEKVFRDVFVAAWTKVMNADRFDV 725

```

```

RESULT 8
US-08-459-499-13
; Sequence 13, Application US/08459499
; Patent No. 5871912
; GENERAL INFORMATION:
; APPLICANT: Heym, Beate
; APPLICANT: Cole, Stewart T.
; APPLICANT: Young, Douglas B.
; APPLICANT: Zhang, Ying
; TITLE OF INVENTION: Nucleic Acid Probes, Sequences, and Methods
; TITLE OF INVENTION: for Detecting Mycobacterium Tuberculosis Resistant to Isonit
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Parabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,499
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/875,940
; FILING DATE: 30-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/929,206
; FILING DATE: 27-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/029,655
; FILING DATE: 11-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0110-03000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 726 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-459-499-13

Query Match 51.6%; Score 2093; DB 2; Length 726;
Best Local Similarity 53.4%; Pred. No. 8e-189;
Matches 405; Conservative 103; Mismatches 206; Indels 44; Gaps 10;

QY 4 ASADLHSSLQRCRAFPVLPSPHRAIRERAMSGKCPVMHGG---NTSTGTSNKDWPE 60
DB 2 STSDDIHT-----TATGKCPFHGGHDQSGAGATTTTRDMPN 39
QY 61 GLNLDTLHQDRKSDPMDPDFNYREEVRKLDLFDALKKDVHMLTDSQEWMPADMGHYGGL 120
DB 40 QLRVDDLNOHNSRNPGLGEDFYRKEFSKLDYGLKDLKALLETESQWPPADMGSGYAGL 99
QY 121 MIRMAHSGAGTYRIADRGGGGTGSGORFAPLNSWPDVSLDKARRLLWPDKKYGKISW 180
DB 100 FIRMAHSGAGTYRISIDRGGAGRGQRFAPLNSWPDVSLDKARRLLWPDKKYGKISW 159
QY 181 ADLMILAGTVAYESMGLPAYGFSRGRVDIWEPEKDIYWGDEKELASDERYGVNKPET 240

```

Db 160 ADLFLAGNVALENSGRTFGAGREDVWEPDLVNMGDRAWLTH-----RHPEA 211
QY 241 M-ENPLAAVOMGLIYVNPGEVNGHPDLRTAQOVLTFARMAMNDKTAALTAGGHTVGN 299
Db 212 LAKAPLGATEMGLIYVNPGEVNGHPDHSGEPLSAAAIIRATFGNMGNDDEETVALLTAGGHTLGK 271
QY 300 CHGNNGASALADPPKASDVENOGLGWGNMGKASNAVTSIEGAWTNTPKFDMGYFD 359
Db 272 THGAGPTS NVGDPPEAAPIEEOGLGWAISTYSGVGADAITSGLEVVTQTPTQWSNYFFE 331
QY 360 LLEGVNWLKSPAGAHWEPIIDIKENKPVDPASDPSIRHNPIMTDADMAIKVNPYRAI 419
Db 332 NLFKEVWOTRSPAGAIQFEAVD-APEIIP-DFFDPSKRRKPTMLVDTLTLRFDPPEFKI 389
QY 420 CEKFMADPEYFKKTFKAWFKLTHRDGLGPKSRYIGPEVPAEDLIWQDPIAGNTDYCEV 479
Db 390 SRRLNDPQAFNEAFARAWFKLTHRDGMPKSRIGPEVPKEDLIWQDPLPQIYNPTEQD 449
QY 480 V---KQIAQSGLSISEMSTANDSARTYRSGDMRGANGARIRLAPQNEWOGNEPERLA 536
Db 450 IIDLKFAIADSGLSVSELVSAWASASTFRGGDKRGANGARLALMPQDMDVN--AAAV 507
QY 537 KVLVYVEQISADTG-ASIADVIVLAGSVGIERAKAAAGYDVRVPLKGRGDATAETDAD 595
Db 508 RALPVLKIKESGRASLADIIVLAGVGVGEKASAAAGLSIHVPFAPGRVDAQDQTDIE 567
QY 596 SFAPLEPLADGFRNMOKKEYVYKPEMLLDRAQLMGLTGPEMTVLGGHVRVLGTYGGTK 655
Db 568 MEELLEPIADGFRNRYARLDVSTESLLIDKAQQLTLTAPENTALVGGHVRVLGGNFDGSK 627
QY 656 HGVFTDCEQLTNDFFVNLTDGNSWKPV--GSNAYEIRDRKTGAVKWTASRVLDLVFGSN 713
Db 628 NGVFTDRVGLVNSDFVNLDMRYEWMKATDESKELFEGREDRETGEYKFTASRADLVFGSN 687
QY 714 SLLRSVAEYVAODDNGEKEFVRDFVAAWTKVMNADREFV 751
Db 688 SVLRVAEYVAODDNGEKEFVRDFVAAWTKVMNADREFD 725

RESULT 9

US-09-082-614A-49
Sequence 49, Application US/09082614A
Patent No. 6124098
GENERAL INFORMATION:
APPLICANT: Reym, Beate
APPLICANT: Cole, Stewart
APPLICANT: Young, Douglas
APPLICANT: Zhang, Ying
APPLICANT: Honore, Nadine
APPLICANT: Telenti, Amalio
APPLICANT: Bodmer, Thomas
TITLE OF INVENTION: Rapid Detection of Antibiotic Resistance
TITLE OF INVENTION: In Mycobacterium Tuberculosis
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Danner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/082,614A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/313,185
FILING DATE: 12-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 02356.0068-00000
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 726 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-082-614A-49
Query Match 51.6%; Score 2093; DB 3; Length 726;
Best Local Similarity 53.4%; Pred. No. 8e-189;
Matches 405; Conservative 103; Mismatches 206; Indels 44; Gaps 10;
QY 4 ASADDLHSSLQQRCAFPVLPSPRHRAIRERAMSGKCPVMHGS---NTSTGT-SNKDWPE 60
Db 2 STSDDIHNT-----TATGKCPHQGHQDSAGAGTTTRDWWPN 39
QY 61 GLNLDILHQDRKSDPMDPDFNYREVRKLDLFDALKKDVHMLTDSQEWPPADWGHYGL 120
Db 40 QLRVLLNQHNSRNPGLGDFDYRKEFSKLDYVGLKDLKALLTQESQPMWADWGSYAGL 99
QY 121 MTRMAHWSAGTRIADGRGGGTGSGORFAPLNSWPDNVSLDKARLLWPIKKYGNKISW 180
Db 100 FIRMAHWCAGTYSIDGRGGAGGQORFAPLNSWPDNVSLDKARLLWPIKKYGNKISW 159
QY 181 ADLMLLAGVAVYESMGLPAYGFSGRVDIWEPEKDIYWGDERKEWLAPSERYGVNKPET 240
Db 160 ADLFLAGNVALENSGRTFGAGREDVWEPDLVNMGDRAWLTH-----RHPEA 211
QY 241 M-ENPLAAVOMGLIYVNPGEVNGHPDLRTAQOVLTFARMAMNDKTAALTAGGHTVGN 299
Db 212 LAKAPLGATEMGLIYVNPGEVNGHPDHSGEPLSAAAIIRATFGNMGNDDEETVALLTAGGHTLGK 271
QY 300 CHGNNGASALADPPKASDVENOGLGWGNMGKASNAVTSIEGAWTNTPKFDMGYFD 359
Db 272 THGAGPTS NVGDPPEAAPIEEOGLGWAISTYSGVGADAITSGLEVVTQTPTQWSNYFFE 331
QY 360 LLEGVNWLKSPAGAHWEPIIDIKENKPVDPASDPSIRHNPIMTDADMAIKVNPYRAI 419
Db 332 NLFKEVWOTRSPAGAIQFEAVD-APEIIP-DFFDPSKRRKPTMLVDTLTLRFDPPEFKI 389
QY 420 CEKFMADPEYFKKTFKAWFKLTHRDGLGPKSRYIGPEVPAEDLIWQDPIAGNTDYCEV 479
Db 390 SRRLNDPQAFNEAFARAWFKLTHRDGMPKSRIGPEVPKEDLIWQDPLPQIYNPTEQD 449
QY 480 V---KQIAQSGLSISEMSTANDSARTYRSGDMRGANGARIRLAPQNEWOGNEPERLA 536
Db 450 IIDLKFAIADSGLSVSELVSAWASASTFRGGDKRGANGARLALMPQDMDVN--AAAV 507
QY 537 KVLVYVEQISADTG-ASIADVIVLAGSVGIERAKAAAGYDVRVPLKGRGDATAETDAD 595
Db 508 RALPVLKIKESGRASLADIIVLAGVGVGEKASAAAGLSIHVPFAPGRVDAQDQTDIE 567
QY 596 SFAPLEPLADGFRNMOKKEYVYKPEMLLDRAQLMGLTGPEMTVLGGHVRVLGTYGGTK 655
Db 568 MEELLEPIADGFRNRYARLDVSTESLLIDKAQQLTLTAPENTALVGGHVRVLGGNFDGSK 627
QY 656 HGVFTDCEQLTNDFFVNLTDGNSWKPV--GSNAYEIRDRKTGAVKWTASRVLDLVFGSN 713
Db 628 NGVFTDRVGLVNSDFVNLDMRYEWMKATDESKELFEGREDRETGEYKFTASRADLVFGSN 687
QY 714 SLLRSVAEYVAODDNGEKEFVRDFVAAWTKVMNADREFV 751
Db 688 SVLRVAEYVAODDNGEKEFVRDFVAAWTKVMNADREFD 725

```

RESULT 10
US-09-328-352-8115
; Sequence 8115, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Broton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 8115
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-8115

```

Query Match	51.1%	Score	2072.5	DB 4	Length	726
Best Local Similarity	53.2%	Pred.	No. 7e-187			
Matches	394	Conservative	106	Mismatches	215	Indels
	25	Gaps	8			
QY	24	VSPRHAIIRERAMSGKPVMHGGTST-----GTSNKDWPEGLINDILHQQRKSDPM	77			
Db	1	ILPRENIMSN--BSKCP-FSGHNSKPOQTVVGGCTANLHWPNQRLDYLHQSERSNPL	57			
QY	78	DPDENREYERKLFDFALKDVHAIITDSDQWNPADMGHYGGLMIRMAWISAGYTRIADG	137			
Db	58	DKDENYRQEFKLDYALKADIKNVLDTSQDWNPADMGNTGTGLFIRLAWHAAGYRMDG	117			
QY	138	RGGGTGSQRPAPLNSPNDVNSDKARLLWPIKKYGNKISWADMLIAGTVAYESMGL	197			
Db	118	RGGAGRQQRPAPLNSPNDNASLKDARLLWPIVVKYQKISWADMLIAGTIALESSGF	177			
QY	198	PAYGFSGRVDIWEPEKDIYWGDEKEWLA PSDERYGVNKPETMENPLAAVQCLIIYNP	257			
Db	178	RTFGGAGREDVWEPDNDVWGDEKEWLAIH-----NSEALAGSNLAATEMGLIIYNP	230			
QY	258	EGVNGHPDPILRTAQVLETFARMAMNDEKTAALTAGGTYGCHGNGSNASALAPDPKASD	317			
Db	231	EGPOASGDRPSAAPPFIRATEFGNAMODEEIVALLAGGHTLGKTHGAAPADHVQADPEGAP	290			
QY	318	VENOGLGWGNPMQGNKASNATVSGIEGATNTNPKFDKMGYFDLLFGYNWELKSPAGAAH	377			
Db	291	IEQMGFWANSYGTGCVGKDAITSLEVIWISOTTQMSNYEFENLFKYEYQERSPAGAIQ	350			
QY	378	WEPIDIKKENKPVADSPSIRHNPIMTDADMAIKVNPTYRAICEKFMADPEYFKKTFKA	437			
Db	351	WVAAD--AEAIIPDPDPSIKRPTMLTTDLTLRFDEPEKISRRFLNDPQAFANAPARA	408			
QY	438	WFKLTHRDLPKSYIGPEVPADSLIWQDPIPAAGNTDYCEEVV---KOKIAQSGLSISEM	494			
Db	409	WFKLTHRDMGPKARYLGPEVPADSLIWQDPLPAASATPPSSASTADAKAKIVALGLSAGEL	468			
QY	495	VSTAWSARTYRGSDMRGGANGARIILAPQNEQNGNEPERLAKVLSYVEIOISADTGASIA	554			
Db	469	VSLWASASTFRGGDKRGGANGARIALSPOREWEVN--KKAVETLLKIEELKRASTQLSLA	526			
QY	555	DVIVLAGSVIEKAAKAAAGYDVVPFLKRGDATAEMTDADSPAPLEPADGFRNNQKKE	614			
Db	527	DLIVLAGNVGEQAQNAAGENIYVPAPGRVDALQSGTDVESQQLLGLGLADGFRNNKKG	586			
QY	615	YVVKPEMLLDRAQLMGLTGPEMTVILGGMRVLYGTNGKTKHGVTFTDCEQLTFNDFVNL	674			
Db	587	VNTPAEVLLIDKAAQQLTAPELTAITGLLGLRVLTGNTWDSOHCVFVTTQOVGVLTDFETNL	646			
QY	675	TDMGNSKKPVGSNA--YELRDRTGAKVTASRVLDVFGSNLLRSYAEVYAQDDNGKEF	732			
Db	647	LDMSNVWAPVDSTSEVFEGDKRGSVTKFTATRNDLVFGSNSTRALAEYIAQADGKEF	706			
QY	733	VRDFVAWTKVMNADRFDA	752			

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	50.4%	Score 2044.5	DB 1	Length 740
Best Local Similarity	54.3%	Pred. No. 3.2e-184		
Matches	406	Conservative 93	Mismatches 220	Indels 29
Gaps	13			
QY	24	VSPRHRAIRER--AMSGKCPVM-HGGNTSTGTSTNKDWPEGLNLDILHQDRKSDPMDP	79	
Db	1	VPQHPPIETTTGAAGNGCPVGHMKYPVEGGNQDWPNNRLNLKVLHQPAVADPMA	60	
QY	80	DFNYREVRKLDFAALDKDHALMTDSQEWPPADWGHYGGILIRMAHWSAGTYRTADCG	139	
Db	61	AFDYAAEVAIDYDALTRDIEVNTTSQPPWPADYGHYGLPIRMAHWAAGTYRIHDG	120	
QY	140	GGGTGSGRFAPLNSWPONSLDKARRLLWPIKKYGNKISWADLMILAGTVAYESMGLPA	199	
Db	121	GAGGGMQRFAPLNSWPONSLDKARRLLWPVKYKGLSWADLIIVPAGNCALES	180	
QY	200	XGFSFGKVDIWEPEKDIYWGDKEWLAPESDERYGDVNPKEPTWENPLAIVQMG	259	
Db	181	FGFGFGKVDQWEPD-EYVWKEATWL--GDERY---SGKRDLENPLAIVQMG	234	
QY	260	VNGHPDPLRTAQOVLETFARMANDEKTAALTAGCHTVGNCHGNNGNSALADPDKASDVE	319	
Db	235	PNGNPDPMAAAVDIRETFRRMANNDVETALIVGGHTFGKTHGAGPADLVCPPEEA	294	
QY	320	NQGLGHGNPNMQGKASNAVTSGIEGATWNTPTFDNGYDFDLLFCYNNMELKKS	379	
Db	295	QMGLGKSSYGTGTGDKDAITSGIEVWVNTPTKWDNSFLEILTYGYEWELTKSPAGAWOYT	354	


```

; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,782
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D.
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 150.141US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 740 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-418-782-7

Query Match      50.3%; Score 2041.5; DB 1; Length 740;
Best Local Similarity 54.4%; Pred. No. 6.1e-184;
Matches 405; Conservative 92; Mismatches 218; Indels 29; Gaps 13;

QY 28 HRAIRER---AMSKCPVM-HGGNTSTGTSNKDWPEGLNLDILHQDRLSDPMDPDFNY 83
Db 5 HPPITETTGAASNGCPVGHMKYPVEGGNQDMPNRLNLKVLHQNPVADPMGAAPDY 64
QY 84 REEYRKLDFOALKKDVHMLTDSQEMWPDWGHYGGMLRMWHSAGTYRIADCRGGGT 143
Db 65 AAEVATIDVADTRDIEEVMTTSQWPWADYGHGPLFIRMWHAAGTYRIHDSRGAGG 124
QY 144 GSQFAPLNSWPONVSLDKARRLLWPIKKYKNKISWADLMILAGTVAYESMGLPAYGFS 203
Db 125 GMORFAPLNSWPONASLDKARRLLWPKYKKYKLSWADLVFAGNCALESMSGFKTFGG 184
QY 204 FGRVDIWEPEKDIYWGDEKEWLAPSDERYGDVKNKPETMENPLAIVOMGLIYVNPENGVNH 263
Db 185 FGRVDQWEPD-EVTWGKEATWL--GDERY---SKRDLENPLAAVQNGLIYVNPENGN 238
QY 264 PDPLRTAQVLETFARMAMDEKTAALTAGGTVGNGHGNASALAPDPKASDVENQGL 323
Db 239 PDPMAAAVDIRETFARMAMDVETAAALIVGGHTFGKTHGAGPADLVGPEPEAPLEQML 298
QY 324 GWGNPNMQKASNAVTSGEGAWTNTPTKFDMGYFDLLFGYNNWELKSPAGAHWEPID- 382
Db 299 GWKSSYGTGTGKDAITSGIEVYVWNTPTKWDNSFLIYGYEWELTKSPAGAWQYTAKG 358
QY 383 IKKENKVPDASDSIRNPNTDADMAIKVNPYRAICEFMDPEYFKTFKAWTKLT 442
Db 359 AGAGTIDPPGGGG--RSPMTLATDSLVDPIYERITRRLWLEHPEADEFAKAWTKLI 416
QY 443 HRDLGPKRSYIGPEVPAEDLIWDDIPAGNTDYCEEV---VKOKIAQSGLSISEMYSTA 498
Db 417 HRDMPVARYLGLPLVPKQTLWQDDPVPAVSHDLVGEAEIASLSQIRASGLTVSOLVSTA 476
QY 499 WDSARTYRSGDMRGGAIRLAPONEWGNPEE-FLAKVLSYEQI-----SADTG-- 550
Db 477 WAAASSPRGSKRGGAIRLQPVQGVENVNDPDGDLRKYVIRITLSEIQESFNSAAPGNI 536
QY 551 -ASTADVILVAGSVGIEKAAAGYDVRVPLFAGRGDATAEMTDADSFAPLEPLADFRN 609
Db 537 KVSFADLVVJGGCAIEKAAAGAHNITVPTFGRTDASQEQTDVESFAVLEPKADGFRN 596
QY 610 WQKYEYVYKPEEMLLDRAQLMGLTGPMTVLLAGMRVLGTYGGTKHGVTFDCEGOLTD 669
Db 597 YLKGKNPLPAEYMLLDKANLLTSLAPMTVLVGLRLVGLGANYKRLPLGVTFEASESLTND 656

```

```

QY 670 FFVNLTDMGNSWK--PVGSNAYEIRDRKTKGAKWTASVDLVFGNSLLRSYAEVYADD 727
Db 657 FFVNLLDNGITWEPSPADDDGTGQKDGSGKVKWTGSRVDLVFGNSLRLALVEYVGADD 715
QY 728 NGEKFRVDFVAWTKVMNADRFV 751
Db 716 AQPKEVODFVAADKVMNLDREFV 739

```

RESULT 14

```

US-08-228-662-7
; Sequence 7, Application US/08228662
; Patent No. 568639
; GENERAL INFORMATION:
; APPLICANT: COCKERILL, FRANKLIN R.
; APPLICANT: KLINE, BRUCE C.
; APPLICANT: UHL, JAMES R.
; TITLE OF INVENTION: DETECTION OF ISONIAZID RESISTANT STRAINS
; TITLE OF INVENTION: OF M. TUBERCULOSIS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCHWEGMAN, LUNDBERG & WOESSNER, P.A.
; STREET: 3500 IDS CENTER
; CITY: MINNEAPOLIS
; STATE: MINNESOTA
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/228,662
; FILING DATE: 18-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WOESSNER, WARREN D.
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 150.123US01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 740 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-228-662-7

```

```

Query Match      50.3%; Score 2041.5; DB 1; Length 740;
Best Local Similarity 54.4%; Pred. No. 6.1e-184;
Matches 405; Conservative 92; Mismatches 218; Indels 29; Gaps 13;

QY 28 HRAIRER---AMSKCPVM-HGGNTSTGTSNKDWPEGLNLDILHQDRLSDPMDPDFNY 83
Db 5 HPPITETTGAASNGCPVGHMKYPVEGGNQDMPNRLNLKVLHQNPVADPMGAAPDY 64
QY 84 REEYRKLDFOALKKDVHMLTDSQEMWPDWGHYGGMLRMWHSAGTYRIADCRGGGT 143
Db 65 AAEVATIDVADTRDIEEVMTTSQWPWADYGHGPLFIRMWHAAGTYRIHDSRGAGG 124
QY 144 GSQFAPLNSWPONVSLDKARRLLWPIKKYKNKISWADLMILAGTVAYESMGLPAYGFS 203
Db 125 GMORFAPLNSWPONASLDKARRLLWPKYKKYKLSWADLVFAGNCALESMSGFKTFGG 184
QY 204 FGRVDIWEPEKDIYWGDEKEWLAPSDERYGDVKNKPETMENPLAIVOMGLIYVNPENGVNH 263
Db 185 FGRVDQWEPD-EVTWGKEATWL--GDERY---SKRDLENPLAAVQNGLIYVNPENGN 238

```

```

QY 264 PDLRTAQQVLETFARMAMNDEKTAALTAGGHTVGNCHGNGNASALAPDPKASDVNQGL 323
Db 239 PDPMAAANDIRETFRRMAMNDVETAAALIVGGHTFKTHGAGPADLVGPEPEAAPLEQMG 298
QY 324 GNGNPNMOCKASNAVTSIGEGAWTNTPTKFDMGYFDLLFGYNWELKKSAPAGAHWEPID- 382
Db 299 GWKSSVGTGTGKDAITSGLEVVTWPTKWDNSFLEILYGEWELTKSPAGAWQYAKDG 358
QY 383 IKENKPVDSIRHNPIMTDADMAIKVNPYRAICEKFMADPEYFKKTFKAWFKLT 442
Db 359 AGAGTIPDPFGGPG--RSPTMLATDLSLRVDPIYERITRRWLEHPEELADEFAKAWYKLI 416
QY 443 HRDLGPKSVYIGPEVPAEDLIWODPIPAAGNTDYCEV---VKQIAQSGLSISEMVSTA 498
Db 417 HRDMGPVARYLGPLVPKQTLWQDPVPVAVSHDLVGEAEIASLKSQIRASGLTVSOLVSTA 476
QY 499 WDSARTYGRSDMRGGANGARIRLAPONEWOGNEPE--RLAKVLSVYEQI-----SADTG-- 550
Db 477 WAAASSFRGSDKRGANGGRIQLQVQGWENVDPPGDLRKVIRTLLEEIOESFNSAAPGNI 536
QY 551 -ASIADVIVLAGSVGIEKAAAGYDVRVPFLKRGDATAEMTDADSFAPLEPLADGFRN 609
Db 537 KVSFADLVVILGGAIEKAAAGHNITVPFTPGRTDASQEQTDVESFAVLEPKADGFRN 596
QY 610 WQKEKVVVKEPEMLDRAQLMGLTGPEMTVLLGGHRYLGTNTYGGTKHGVTTCGQLTND 669
Db 597 YLGKGNPLPAEYMLLDKANLLTSAPEMTVVLGGRLVGLGANYKRLPLGVFTFSESLTND 656
QY 670 FVNLTDGMSNK--PVGSNAVEIRDRKTGAVKWTASRYDLVFGSNLSLSYAEVYQADD 727
Db 657 FVNLDDMGITWEPSADGTYQKGD--GSGKVKWTGSRVDLVFGSNLSRLVYEVYQADD 715
QY 728 NGEKVFVRDFAAWTKVMNADRDV 751
Db 716 AOPKFVQDEVAADKVMNLDREVD 739

```

RESULT 15

```

US-08-852-219-7
; Sequence 7, Application US/08852219
; Patent No. 5922575
; GENERAL INFORMATION:
; APPLICANT: Cockerill, Franklin R.
; APPLICANT: Kline, Bruce C.
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Isoniazid Resistant Strains
; TITLE OF INVENTION: Of M. Tuberculosis
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mueeting, Raasch & Gebhardt, P.A.
; STREET: 119 No. 5922575th Fourth Street, Ste. 203
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,219
; FILING DATE: 07-May-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sandberg, Victoria A.
; REGISTRATION NUMBER: 41,287
; REFERENCE/DOCKET NUMBER: 230.00010130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1226
; TELEFAX: 612-305-1228
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 740 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-852-219-7

Query Match      50.3%; Score 2041.5; DB 2; Length 740;
Best Local Similarity 54.4%; Pred. No. 6.1e-184;
Matches 405; Conservative 92; Mismatches 218; Indels 29; Gaps 13;

QY 28 HRAIRER---AMSGKCPVM-HGNTSTGTSTNDKDWPEGLNLDLHQDQRKSDPMDPDFNY 83
Db 5 HPIITETTTGAASNGCPVVGHMKYPVGEGGNODWPNRLNLKVLHQNPADVPMGAADF 64
QY 84 REVRRLDKDALKDVHMLTDSQEWPPADWGHYGLMIRMAWHSAGTTRIADRGGGGT 143
Db 65 AAEVATIDYDALTRDIEEVTTSQPMWADYGHYGLFIRMAWHAAGTTRIHDGCGAGG 124
QY 144 GSQRFAPLNSPDVNSLDKARLLWPICKKYGKISWADLMILAGTVAYESMGLPAYGS 203
Db 125 GMORFAPLNSPDNASLDKARLLWPVKKYKKLSWADLIVFAGNCALESMEGFKTFGG 184
QY 204 FGRVDIWEPEKDIYWCDEKWLAPSDERYGDVYKPTMENPLAAVOMGLIYVPEGVNH 263
Db 185 FGRVDQWEPD-EYNGKEATWL--GDERY---SGKRDLENPLAAVOMGLIYVPEGPN 238
QY 264 PDLRTAQQVLETFARMAMNDEKTAALTAGGHTVGNCHGNGNASALAPDPKASDVNQGL 323
Db 239 PDPMAAANDIRETFRRMAMNDVETAAALIVGGHTFKTHGAGPADLVGPEPEAAPLEQ 298
QY 324 GNGNPNMOCKASNAVTSIGEGAWTNTPTKFDMGYFDLLFGYNWELKKSAPAGAHWEPID- 382
Db 299 GWKSSVGTGTGKDAITSGLEVVTWPTKWDNSFLEILYGEWELTKSPAGAWQYAKDG 358
QY 383 IKENKPVDSIRHNPIMTDADMAIKVNPYRAICEKFMADPEYFKKTFKAWFKLT 442
Db 359 AGAGTIPDPFGGPG--RSPTMLATDLSLRVDPIYERITRRWLEHPEELADEFAKAWYKLI 416
QY 443 HRDLGPKSVYIGPEVPAEDLIWODPIPAAGNTDYCEV---VKQIAQSGLSISEMVSTA 498
Db 417 HRDMGPVARYLGPLVPKQTLWQDPVPVAVSHDLVGEAEIASLKSQIRASGLTVSOLVSTA 476
QY 499 WDSARTYGRSDMRGGANGARIRLAPONEWOGNEPE--RLAKVLSVYEQI-----SADTG-- 550
Db 477 WAAASSFRGSDKRGANGGRIQLQVQGWENVDPPGDLRKVIRTLLEEIOESFNSAAPGNI 536
QY 551 -ASIADVIVLAGSVGIEKAAAGYDVRVPFLKRGDATAEMTDADSFAPLEPLADGFRN 609
Db 537 KVSFADLVVILGGAIEKAAAGHNITVPFTPGRTDASQEQTDVESFAVLEPKADGFRN 596
QY 610 WQKEKVVVKEPEMLDRAQLMGLTGPEMTVLLGGHRYLGTNTYGGTKHGVTTCGQLTND 669
Db 597 YLGKGNPLPAEYMLLDKANLLTSAPEMTVVLGGRLVGLGANYKRLPLGVFTFSESLTND 656
QY 670 FVNLTDGMSNK--PVGSNAVEIRDRKTGAVKWTASRYDLVFGSNLSLSYAEVYQADD 727
Db 657 FVNLDDMGITWEPSADGTYQKGD--GSGKVKWTGSRVDLVFGSNLSRLVYEVYQADD 715
QY 728 NGEKVFVRDFAAWTKVMNADRDV 751
Db 716 AOPKFVQDEVAADKVMNLDREVD 739

```

Search completed: October 7, 2003, 19:47:34
Job time : 26.1175 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 7, 2003, 19:45:48 ; Search time 48.759 Seconds
(without alignments)
2443.330 Million cell updates/sec

Title: US-09-884-889-6
Perfect score: 4058
Sequence: 1 MNASADDLHSSLQQRCAF.....RDFVAAWTKVMNADRFDVAS 753

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 587654 seqs, 158212981 residues

Total number of hits satisfying chosen parameters: 587654

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1:	/cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2:	/cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3:	/cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4:	/cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5:	/cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6:	/cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7:	/cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8:	/cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9:	/cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10:	/cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11:	/cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12:	/cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13:	/cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14:	/cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15:	/cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16:	/cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17:	/cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18:	/cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4058	100.0	753	10	US-09-884-889-6
2	2290.5	56.4	745	10	US-09-884-889-8
3	272.5	6.7	291	10	US-09-870-501-1
4	272	6.7	250	14	US-10-027-559-6
5	234.5	5.8	287	12	US-10-342-224-18
6	167.5	4.1	176	10	US-09-734-017A-54
7	123	3.0	1036	12	US-10-259-165-180
8	123	3.0	4471	15	US-10-205-032-10
9	120	3.0	357	14	US-10-027-559-12
10	117.5	2.9	4881	15	US-10-156-761-8481
11	115	2.8	3613	15	US-10-156-761-10432
12	115	2.8	4551	11	US-09-793-708-1
13	115	2.8	4551	12	US-10-201-365-2
14	115	2.8	4551	12	US-10-160-539-1
15	115	2.8	4551	10	US-09-861-289-31

ALIGNMENTS

RESULT 1

US-09-884-889-6
; Sequence 6, Application US/09884889
; Patent No. US20020102680A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: ROBERTSON, Dan
; APPLICANT: ADHIKARI, Robert
; APPLICANT: SANYAL, Indrajit
; TITLE OF INVENTION: CATALASES
; FILE REFERENCE: DIVER1100-4
; CURRENT APPLICATION NUMBER: US/09/884,889
; CURRENT FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 09/412,347
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: US 08/951,844
; PRIOR FILING DATE: 1997-10-16
; PRIOR APPLICATION NUMBER: US 08/674,887
; PRIOR FILING DATE: 1996-07-03
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 753
; TYPE: PRT
; ORGANISM: Alcaligenes (Deleya) aquamarinus
US-09-884-889-6

Query Match 100.0%; Score 4058; DB 10; Length 753;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 753; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNASADDLHSSLQQRCAFVPLVSPRIAIRERAMSGKCPVMHGGNTSTGTSNKDWPE 60
Db 1 MNASADDLHSSLQQRCAFVPLVSPRIAIRERAMSGKCPVMHGGNTSTGTSNKDWPE 60
QY 61 GLNLDILHQDQRKSDPMDPDNFNYREEVRKLDLFDALKDVKHMTDSEWNPADGHGGL 120
Db 61 GLNLDILHQDQRKSDPMDPDNFNYREEVRKLDLFDALKDVKHMTDSEWNPADGHGGL 120

Sequence 31, Appl
Sequence 31, Appl
Sequence 31, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 381, App
Sequence 4303, Ap
Sequence 10906, A
Sequence 8, Appl
Sequence 4, Appl
Sequence 90, Appl
Sequence 2, Appl
Sequence 12458, A
Sequence 5455, Ap
Sequence 9246, Ap
Sequence 4, Appl
Sequence 113, App
Sequence 8616, Ap
Sequence 8762, Ap
Sequence 11828, A
Sequence 2, Appl
Sequence 4495, Ap
Sequence 4899, Ap
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl

121 MTRMAHWSAGTYRIADGRGGGTGSGORFAPLNSWPNVSLDKARLLWPPIKKYGNKISW 180
 121 MTRMAHWSAGTYRIADGRGGGTGSGORFAPLNSWPNVSLDKARLLWPPIKKYGNKISW 180
 181 ADLMLAGTAYESGLPAYGSGRVDIWEPEKDIYWGDEKWLAPSDERYGDNKPT 240
 181 ADLMLAGTAYESGLPAYGSGRVDIWEPEKDIYWGDEKWLAPSDERYGDNKPT 240
 241 MENPLAAVOMGLIYVPEGVNCHPDLRTAQOVLTFARMANDEKTAALTAGGHTVNC 300
 241 MENPLAAVOMGLIYVPEGVNCHPDLRTAQOVLTFARMANDEKTAALTAGGHTVNC 300
 301 HGNASALAPPKASDVENOGSLGNGNPNMQKASNAVTSIGEGAWTNTPTKFDGYFDL 360
 301 HGNASALAPPKASDVENOGSLGNGNPNMQKASNAVTSIGEGAWTNTPTKFDGYFDL 360
 361 LFYGNWELKSPAGAHHEPIDIKKENKPVDSIRHNPMTDADMAIKVNPYRATC 420
 361 LFYGNWELKSPAGAHHEPIDIKKENKPVDSIRHNPMTDADMAIKVNPYRATC 420
 421 EKFMADPEYFKKTFKAFKFLTHRDLPKRSRYIGPEVPAEDLIWQDPIPAAGNTDYCEV 480
 421 EKFMADPEYFKKTFKAFKFLTHRDLPKRSRYIGPEVPAEDLIWQDPIPAAGNTDYCEV 480
 481 KOKIAQSGLSISEMSTAWSDARTYRGSDMRGANGARIRLAPQNEWQNEPERLAKVLS 540
 481 KOKIAQSGLSISEMSTAWSDARTYRGSDMRGANGARIRLAPQNEWQNEPERLAKVLS 540
 541 VYEQISADTGASIDVIVLAGSVGTEKAAKAGYDVRVPLFKRGDGTAEWTDADSFAPL 600
 541 VYEQISADTGASIDVIVLAGSVGTEKAAKAGYDVRVPLFKRGDGTAEWTDADSFAPL 600
 601 EPLADGFRNWKQKEYVKAPEEMLLDRAQLMGLTGPEMVLGCMRVLTNYGTHGVPT 660
 601 EPLADGFRNWKQKEYVKAPEEMLLDRAQLMGLTGPEMVLGCMRVLTNYGTHGVPT 660
 661 DCEGQLTNDFFVNLDMGNSKVPVGSNAYEIRDRKTGAVKWTASRVDLVFGSNLLRSYA 720
 661 DCEGQLTNDFFVNLDMGNSKVPVGSNAYEIRDRKTGAVKWTASRVDLVFGSNLLRSYA 720
 721 EYVAQDDNGEKFVRDVFVAWTKVMNADRFDVAS 753
 721 EYVAQDDNGEKFVRDVFVAWTKVMNADRFDVAS 753

RESULT 2

US-09-884-889-8
 ; Sequence 8, Application US/09884889
 ; Patent No. US20020102680A1
 ; GENERAL INFORMATION:
 ; APPLICANT: DIVERSA CORPORATION
 ; APPLICANT: ROBERTSON, Dan
 ; APPLICANT: SANYAL, Indrajit
 ; APPLICANT: ADHIKARI, Robert
 ; TITLE OF INVENTION: CATALASES
 ; FILE REFERENCE: DIVER1100-4
 ; CURRENT APPLICATION NUMBER: US/09/884,889
 ; PRIOR FILING DATE: 2001-06-19
 ; PRIOR APPLICATION NUMBER: US 09/412,347
 ; PRIOR FILING DATE: 1999-10-05
 ; PRIOR APPLICATION NUMBER: US 08/951,844
 ; PRIOR FILING DATE: 1997-10-16
 ; PRIOR APPLICATION NUMBER: US 08/674,887
 ; PRIOR FILING DATE: 1996-07-03
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 8
 ; LENGTH: 745
 ; TYPE: PRT
 ; ORGANISM: Microscilla furvescens
 US-09-884-889-8

Query Match 56.4%; Score 2290.5; DB 10; Length 745;
 Best Local Similarity 59.1%; Pred. No. 66-207;
 Matches 433; Conservative 92; Mismatches 183; Indels 25; Gaps 8;
 QY 38 GKCPTVHGG---NTSGTSNKNWPEGLNLDJLHQQDRKSDPMDPDPNRYEEVRKLDFA 94
 DB 18 GKCPTVHGG---NTSGTSNKNWPEGLNLDJLHQQDRKSDPMDPDPNRYEEVRKLDFA 94
 QY 95 LKDDHALMTDQSEWNPADWYHGLMIRMAHWSAGTYRIADGRGGGTGSGORFAPLNSW 154
 DB 78 VKADLAALMTDQSEWNPADWYHGLMIRMAHWSAGTYRIADGRGGGTGSGORFAPLNSW 137
 QY 155 PDNVSLDKARRLLWPIKKYGNKISWADLMLAGTAYESGLPAYGSGRVDIWEPEK 214
 DB 138 PDNVSLDKARRLLWPIKKYGNKISWADLMLAGTAYESGLPAYGSGRVDIWEPEK 197
 QY 215 DIYWGDEREWLAPSDERY-GDVNKPETMENPLAAVOMGLIYVPEGVNCHPDLRTAQOV 273
 DB 198 DVYWGAEETWLGDKRYEGD---RELENPLGAVOMGLIYVPEGVNCHPDLRTAQOV 251
 QY 274 LETFARMANDEKTAALTAGGHTVNCNCHGNASALAPDPKASDVENOGSLGNGNPNMQ 332
 DB 252 RETFARMANDEKTAALTAGGHTVNCNCHGNASALAPDPKASDVENOGSLGNGNPNMQ 311
 QY 333 KASNAVTSIGEGAWTNTPTKFDGYFDLFGYNWELKSPAGAHHEPIDIKKENKPVDA 392
 DB 312 HCATITSGLEGAWTKTPTQWSNPNFENLFGYEWELTKSPAGAYQWKPKDGAGATIPDA 371
 QY 393 SDPSIRHNPIMTDADMAIKVNPYRATCEKFMADPEYFKKTFKAWFKLTHRDLPKRSY 452
 DB 372 HDPSKSHAPFMTLTDLALRMDPDYEKISRYYENPDEFADAFKAWYKLTTHRDMPKRY 431
 QY 453 IGPEVPAEDLIWQDPIPAAGNTDYCE---EVYKQKTAOSGLSISEMSTAWSDARTYGS 508
 DB 432 LGPEVPAEDLIWQDPIPAAGNTDYCE---EVYKQKTAOSGLSISEMSTAWSDARTYGS 491
 QY 509 DMRGANGARIRLAPQNEWQNEPERLAKVLSVYEQISAD-----TGASIDVIVLA 560
 DB 492 DKRGANGARIRLAPQNEWQNEPERLAKVLSVYEQISAD-----TGASIDVIVLA 551
 QY 561 GSVGTEKAAKAGYDVRVPLFKRGDGTAEWTDADSFAPLEADGFRNWKQKEYVVRPE 620
 DB 552 GCAGVEKAAKAGYDVRVPLFKRGDGTAEWTDADSFAPLEADGFRNWKQKEYVVRPE 611
 QY 621 EMLDRAQLMGLTGPEMVLGCMRVLTNYGTHGVPTDCEGQLTNDFFVNLDMGNS 680
 DB 612 EMLDRAQLMGLTGPEMVLGCMRVLTNYGTHGVPTDCEGQLTNDFFVNLDMGNS 671
 QY 681 WK--PVGSNAYEIRDRKTGAVKWTASRVDLVFGSNLLRSYAETVAQDDNGEKFVRDFA 738
 DB 672 WRASDESKVFEGRDFTGEVKNWGTGTRVDLIFGNSSELALAEVYGCADSEKFKVDFVK 731
 QY 739 AWTVMNADRFDV 751
 DB 732 AWTVMNADRFDV 744

RESULT 3

US-09-870-501-1
 ; Sequence 1, Application US/09870501
 ; Patent No. US20020100081A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Takabe, Tetsuko
 ; TITLE OF INVENTION: Peroxisomal Ascorbate Peroxidase Gene Induced by High
 ; TITLE OF INVENTION: Temperature Stress and a Transgenic Plant Exhibiting
 ; TITLE OF INVENTION: Thermotolerance
 ; FILE REFERENCE: 026350-053
 ; CURRENT APPLICATION NUMBER: US/09/870,501
 ; PRIOR FILING DATE: 2001-06-01
 ; PRIOR APPLICATION NUMBER: JP 2000-172850
 ; PRIOR FILING DATE: 2000-06-09
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: PatentIn version 3.1

```
; SEQ ID NO 1
; LENGTH: 291
; TYPE: PRT
; ORGANISM: Hordeum vulgare L. Haruna-niyo
US-09-870-501-1

Query Match
Best Local Similarity 6.7%; Score 272.5; DB 10; Length 291;
Matches 93; Conservative 36; Mismatches 96; Indels 119; Gaps 10;

QY 120 LMIRMAHWSAGTYRIADGGGGTGSORFAPLNSWPDNVSUDKARRLLWPIKKYGNKIS 179
Db :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||:
33 IMRLAHWDAGTYDV-NRTGGANGSIRYEEYTHGNSAGLKIALDLLEPIAKAH-PKIT 90
QY 180 WADLMILAGTVAYESMGLPAYGFSGRVDIWEPEKDIYWGDEKWLAPSDERYGDVNKPE 239
Db :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||:
91 YADLHOLAGVAVVETGPTVEFTIPGRD-----SSVCPREGRL----- 129
QY 240 TMENPLAAVQMLIYVNPGEVNGHPDPLRTAQOVLTFARMAMNDEKTAALTAGHTVGN 299
Db :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||:
130 -----PDAKKGAPHLRDIYFRMGLTDKIDVAL-SGGHSLGK 164
QY 300 CHGNGNASALADPKASDVENOGILGWGNPNMOGKASNAVTSIGIEGAWTNTPTKDMGYF- 358
Db :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||:
165 AH-----PE----- 189
QY 359 DLLFGYNWELKKSPAGAHWEPIDIKKENKPVASDPSIRHNPIMTDADMAIKVNPYRA 418
Db :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||:
190 ELLKGESGLLKL-----TDKALLDDPEFR 216
QY 419 ICEKFMADPEYFKTFKAKWKLHRLDGKSRVIGPEVPAEDL 462
Db :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||:
217 YVELYAKDEDFKDYAESHKLS--ELGFTPRSSGPASTRSDV 258

RESULT 4
US-10-027-559-6
; Sequence 6, Application US/10027559
; Publication No. US20020144307A1
; GENERAL INFORMATION:
; APPLICANT: Simmons, Carl
; TITLE OF INVENTION: Maize Defense-Inducible Genes and Their
; FILE REFERENCE: 035718/239835
; CURRENT APPLICATION NUMBER: US/10/027,559
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 60/243,120
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Zea mays
US-10-027-559-6

Query Match
Best Local Similarity 6.7%; Score 272; DB 14; Length 250;
Matches 96; Conservative 36; Mismatches 102; Indels 132; Gaps 11;

QY 83 YREEVKRLDFDALKKDVHMTDSQEWPPADWGHYGGGLMIRMAHWSAGTYRIADGGGG 142
Db :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||:
12 YSEAV-----DKARRKRLAIEKS-----CAPLMRLAHWSAGTFDVS-SRTGGP 56
QY 143 TGSORFAPLNSWPDNVSUDKARRLLWPIKKYGNKISWADLMILAGTVAYESMGLPAYGF 202
Db :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||:
57 FGTMKHQSOLAHGANAGLDIAVRLLEPIKEEF-PILSYADFYQLAGVAVVETGTPPELFF 115
QY 203 SFGSRVDIWEPEKDIYWGDEKWLAPSDERYGDVNKPEMENPLAAVQMLIYVNPGEVNG 262
Db :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||:
116 HPREDKPPQ-----PPEGRLPDATKGSDBHLRQVFGKQML-----S 152
QY 263 HPDPLRTAQOVLTFARMAMNDEKTAALTAGHTVGNCHGNASALADPKASDVENOG 322

; SEQ ID NO 153
; LENGTH: 447
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-342-224-18

Query Match
Best Local Similarity 5.8%; Score 234.5; DB 12; Length 287;
Matches 89; Conservative 45; Mismatches 106; Indels 133; Gaps 12;

QY 80 DFNTREEVKRLDFDALKKDVHMTDSQEWPPADWGHYGGGLMIRMAHWSAGTYRIADGG 139
Db :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||:
7 DAELYLKEITK-----ARRELKSLIANK-----NCAPIMRLAHWDAGTYD-AQSKT 51
QY 140 GGGTGSORFAPLNSWPDNVSUDKARRLLWPIKKYGNKISWADLMILAGTVAYESMGLPA 199
Db :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||:
52 GGPNGSIRNEEETHGANGSLKIALDLCEGVAKAH-PKITADLQLAGVAVVETGGPD 110
QY 200 YGFSFGRVDIWEPEKDIYWGDEKWLAPSDERYGDVNKPEMENPLAAVQMLIYVNPGE 259
Db :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||:
111 IVFVPGRKD-----SNVCPKEGRL----- 129
QY 260 VNGHPDPLRTAQOVLTFARMAMNDEKTAALTAGHTVGNCHGNASALADPKASDV 319
Db :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||:
130 -----PDAKQGFQHLRDFYRMGLSDKIDVAL-SGGHTLGRAH-----PE----- 168
QY 320 NQGLGWGNPNMOGKASNAVTSIGIEGAWTNTPTKDMGYF--DLLFGYNWELKKSPAGAH 377
Db :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||:
169 -----RSFGDGPWTQEPINFDNSYFVRELLKGESEGLLKL----- 204
QY 378 WEPIDIKKENKPVASDPSIRHNPIMTDADMAIKVNPYRAICEKFMADPEYFKTFKAK 437
Db :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||:
205 -----TDKILLEDPEFRRLVELY-ADEDAFFRDYAES 235
QY 438 WFKLTHRLDGPKS 450
Db :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||:
236 HKLSELGFNPNS 248

RESULT 6
US-09-734-017A-54
```

```

; Sequence 54, Application US/09734017A
; Patent No. US20020142422A1
; GENERAL INFORMATION:
; APPLICANT: Lerchl, Jens
; APPLICANT: Renz, Andreas
; APPLICANT: Ehrhardt, Thomas
; APPLICANT: Reindl, Andreas
; APPLICANT: Cirpus, Petra
; APPLICANT: Bischoff, Friedrich
; APPLICANT: Frank, Markus
; APPLICANT: Freund, Annette
; APPLICANT: Duwenig, Elke
; APPLICANT: Schmidt, Ralf-Michael
; TITLE OF INVENTION: Moss genes from Physcomitrella patens encoding proteins involved
; in the synthesis of amino acids, vitamins, cofactors, nucleotides and
; nucleosides
; TITLE OF INVENTION: the
; TITLE OF INVENTION: synthesis of amino acids, vitamins, cofactors, nucleotides and
; nucleosides
; FILE REFERENCE: BASF-NAE-1331-99-US
; CURRENT APPLICATION NUMBER: US/09734,017A
; CURRENT FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: 60/171,100
; PRIOR FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn Ver. 2.1/WordPerfect
; SEQ ID NO 54
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Physcomitrella patens
; US-09-734-017A-54

```

Query Match	4.1%	Score 167.5;	DB 10;	Length 176;
Best Local Similarity	29.1%;	Pred. No. 1.7e-07;		
Matches	53;	Conservative 21;	Mismatches 61;	Indels 47; Gaps 6;
QY	120	LMIRMAHWSAGTYRIADGRGGGTGSGREAPLNSWPDNVSLDKARRLLWP1KKKYGNKIS	179	
Db	35	IIURLAHWGSGTYD-QESK7GGPLGTIRFGQELAHGANAGLDIAVNL1QPIKEQF-PELS	92	
QY	180	WADLMLTAGVAYESMGLPAFYGSFGVDIWEPEKDIYWGDEKWLAPSDERYGVGNKPE	239	
Db	93	YADFYTLAGVAVVEVTGGPTIPFHPGRKD-----HETCPVEGRLPDATK--	136	
QY	240	TMENPLAAVQMGLIYVNPVGNGHPDPLRTAAQQVLETFARMAMNDEKTAALTACGHTVGN	299	
Db	137	-----GL-----DHLRC-----VFTKMGSLTDKDIIVVLSGAHTLGR	167	
QY	300	CH 301		
Db	168	CH 169		

RESULT 7
 US-10-259-165-180
 ; Sequence 180, Application US/10259165
 ; Publication No. US20030135888A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Zhu, Tong
 ; APPLICANT: Wang, Xun
 ; APPLICANT: Chang, Hur-song
 ; APPLICANT: Briggs, Steven P.
 ; APPLICANT: Cooper, Bret
 ; APPLICANT: Glazebrook, Jane
 ; APPLICANT: Goff, Stephen A.
 ; APPLICANT: Katagiri, Fumiyuki
 ; APPLICANT: Kreps, Joel
 ; APPLICANT: Moughamer, Todd
 ; APPLICANT: Provart, Nicholas
 ; APPLICANT: Ricke, Darrell
 ; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING
 ; FILE REFERENCE: 70030-NP
 ; CURRENT APPLICATION NUMBER: US/10/259,165
 ; CURRENT FILING DATE: 2002-09-26
 ; PRIOR APPLICATION NUMBER: US 60/370,620

```

; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/368,327
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 782
; SOFTWARE: PatentList.Pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 180
; LENGTH: 1036
; TYPE: prt
; ORGANISM: Oryza sativa
US-10-259-165-180

```

Query Match	3.08;	Score 123;	DB 12;	Length 1036;
Best Local Similarity	18.3%;	Pred. No. 0.044;		
Matches 139;	Conservative	99;	Mismatches 274;	Indels 248; Gaps 33;
QY	46	GNTSTGTSNKDWPEGLN--LDILHODRKSDPMDPDNFYREEYRKLDLFALKKDVHALM	103	
Db	92	GLTSVGLSNSESPDSVTQISCLNNAD-----ERVEEL--TLK--NCIS	132	
QY	104	TDSEWNPAD-----NGHYGGLMIRMAWHS---AGTYRIADRGGGGTGS	145	
Db	133	SDAQHEVSAGGSTSGEKPVTMRGLMGNF---TRMSWRASDVANREKLAANR--GDVAN	186	
QY	146	QREAPLNSWPDNYSLOKARLLPIKKYCNKTSWADLMLAGTVAYESHGLPAYGSFSG	205	
Db	187	LRVGD-----PIRENLA--VSFGNNMISRNDASNKMG-----SHG	222	
QY	206	RVDIWEPEKDIYWGDEKEWLAPSDERYGDYKNKPTMETENPLAAV-----QMGLIYVNP	257	
Db	223	--DHANNEFNLPFGNQOPFLSP--RPNQNEQVRVERENALIVSFSARILDQHRSKNVP	277	
QY	258	EGVNGHP-DPLRTAQOVLETFARMAMNDEKTAALTAGGHTVNGCHNGNASALAPDPKA	315	
Db	278	SSGVQSPFPKSVLKGKGVVYQGAEEIQVQ-----GNARTRAPMDKI	319	
QY	316	SDVEN--OCLGNGNPNMOGKASNAVTSGTGEA--WTTNPTKFDGCGYFDLLFCYKNWELK	369	
Db	320	RKIPNIPQDSKARMDDGTIFGSGGNVLPEQCEGSLRELIPAKQTMSKFKEM-----	371	
QY	370	KSPAGAHWEPI-DIKENKPNVDASDPSIRHNPIMTDADMAIKVNTPYTAICEKFMADPE	428	
Db	372	-----HFFKQIILDVDKSHAQGSFLQHLRPSYFTISASNOVKVYIGSYGT-----	415	
QY	429	YFKKTKAKWFLTHRDLPKPSKYIGPEVPAEDLIWQDPTPAGNTDYCEVFWQKIAQSG	488	
Db	416	-----ODLSAPSKLDIATDDIFWTKRYLPKPVESQD-----SNGDNASTIKYQKVGEOG	464	
QY	489	LSISMSVSTAMDSARTYSGDMRGANGARIRLAPONEWQGN-----	530	
Db	465	SIAVRRPVHTF--ANHRGGNSEGVD-----PGALMOGHSSTCYRERPKAABPFYG	514	
QY	531	-----EPERLAKVLSVYEQISADTGCASIADEVILA	560	
Db	515	GSMYPAAQRPSSSGNQSVFELRMLEESWYRSPEISOLKGIPLSNIVSLGVLLFEISVIA	574	
QY	561	GSVGI-----EKAATAAGYDVVPFLKRGED-----ATAEMTADSEAPLEIADGFRNW	610	
Db	575	FCQPIFFQVKVLKLVSVFGYSIQIPLGLRQDNKTPVAVNEEDTES-----GLLGLFSLQ	639	
QY	611	QRKEYVVKPPEMLLDRAQL-----MGLTPGEMTVLLGMRVILGPNY-----G	652	
Db	630	LAEEREMHAUKLSADLASLETDAEVEKRHSRMRGFSLSDMDVLASNDLSGASACALGG	689	
QY	653	GTKHGVTTCBEGQLTNDFFVNLTDMGNSWKPVGSNAYEIR	692	
Db	690	ASLSGPTID-----TSEANIIRKRVNDNALRVR	716	

RESULT 8
US-10-205-032-10
; Sequence 10, Application US/10205032

; PRIOR APPLICATION NUMBER: JP 2001-272697
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10432
; LENGTH: 3613
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8481

Query Match 2.9%; Score 117.5; DB 15; Length 4881;
Best Local Similarity 20.8%; Pred. No. 1.7;
Matches 147; Conservative 72; Mismatches 282; Indels 205; Gaps 36;
QY 58 WPEGINDILHQQRKSDPDPDENYREVRKLFDFDALKVDVHALMTDSQEWNPADWGHY 117
DB 3603 WPAAL-----PE---RDRARHLDPTVAFDHH-----RYWVDTSAGHP 3637
QY 118 GGLMRMAWSAGTYRIADGRGGGTGSGORFAPLNSWPNVSLDKARLLWPIKKYGNK 177
DB 3638 GDL-----SAGLGTAAGPLGGS---AVALAESQELL-----FTGR 3670
QY 178 ISWADLMLTAGTVAYESMGLPAYGFSFGRVDIWEPEKDIYWGDEKEWLAPSDERYGDVVK 237
DB 3671 LSLRTHPLADHAIFGTVLLPGTAIL-----ELAVRAGDEV-----CGTVEE 3713
QY 238 PETWENPLAAVOMGLIYV-----PEGVNGHPDPLRTAQVLETFARMAMDEKTAALTA 292
DB 3714 -LTLRTPLVLPEOGSVILQLSVGAPQGPQTPPEEP-----RTFALYAREDDGLSSSA 3766
QY 293 GG-HTVGNCHGNGNASALA-----PDKASDVE-----NQLGNGNPMQ 331
DB 3767 AATGTEWCHATGVTGTARPAEHTQEPWPADAAPVDDLGWTEYQAGAGLGYG-PVFO 3825
QY 332 GK-----ASNVTSGIEG-----AWTNTPTKFMGYFDLLFGYNWELKKSPAG 374
DB 3826 GLREVWRRGDEVFVAVTLPESTEGQAADAARYALHPALLDAALHPVVLRHEDG---AAAD 3882
QY 375 AHWEPTDIKKENFPVADSPSIKH-----NPMTDADMAIKVNP-----TYRAI 419
DB 3883 GHGWLPL-FSWTGVTVAAAGSASTLHVRLTVRTDEDVGLLATDASGRIVISAGSLAFRPV 3940
QY 420 CEKEM--ADPEYFKKTFKAKFKLTHRD-----LGPKSR----- 451
DB 3941 SAEQLQAARTGYHDLHFIERHPL-HLPTTPARTADWALIGCGARRTAANLERNASWQA 3999
QY 452 YIGPEVPAEDLIWQDPIPAAGNTDYCEEVVKQKIAQSGLSISEMVTAWDSARTYRGSDMR 511
DB 4000 YPDPAALAEALAGAPAPGMVVISCEPDGASAPDTSALTDSAPAGSAPADSTALA 4059
QY 512 GGANGARIR-LAPONEHQNEPERLAK---VLSYVEQISADTGASIAADVIVLAGSVGIEK 567
DB 4060 DATQQAATRVLLALQEWAD---ERLAAACRLALLTHGSVTATPDPEVSD-LAHAAVWGLVR 4116
QY 568 AKAAGYDVRVPFLKRGDGAETAEMTADSFAPLEPLADG-----FRNW-----QKKEVY 616
DB 4117 SVQTEPND---RFL-----LATDDDDASNNALPLAGEPQIALRNGAVIPRTRVPV 4167
QY 617 VKPEMLDLRAQLMGLGPMTVLL-GGMRVLG---TNYGTRKHV 658
DB 4168 RQOPSTTDAD-----WDPEATVLTGTTGLVRLVARHLATAHGV 4208

RESULT 11
US-10-156-761-10432
; Sequence 10432, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10432
; LENGTH: 3613
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10432
Query Match 2.8%; Score 115; DB 15; Length 3613;
Best Local Similarity 22.1%; Pred. No. 1.8;
Matches 136; Conservative 66; Mismatches 199; Indels 214; Gaps 36;
QY 126 WHSAGTYRIADGRGGGTGSGORFAPLNSWPNVSLDKARLLWPIKKYGNKISWADLMI 185
DB 1187 WNPHTVLTITGTTGGIG-----AHLARW---LAKEGAERLL--LVSRSGEQAGS----- 1230
QY 186 LAGTVAYESMGLPAYGFSFGRVDIWEPEKDIYWGDEKEWLAPSD-ERYGDVKNKPTWENP 244
DB 1231 -AAELATLGLGIGA-EVTFACDV-----SDRDALGHVVIAGIPAEHP 1270
QY 245 LAAVOMGLIYVNPPEGVNGHPDPLRTA-----QQVLETFARMAMN-DEKTAALT----- 291
DB 1271 LTAV-----FHTAGVSGYAE-LATATPEHYETVLSAKTRGNRLDALTAELTLEAFVLF 1323
QY 292 AGGHTVGNCHGNGNASALAPDPKASDVENOCILGNPNMCKASNAVTSIGEGAWTTNPT 351
DB 1324 SSGAAVWGSAGTG---AYA---AANAYLDGLAW-NRRARGLVATSVS---WGWKATGM 1372
QY 352 KFDMGYFDLLFGYNWELKKSPAGAHWEP-IDIKENKPVDSADSPSIRHNPIMTDADMAI 410
DB 1373 AAD-----GTEQLARRGVRAPEPALAIKALRQALRQDETAL-----TVTDMDAW- 1417
QY 411 KVNPTY-----RAICEKFMADPE-----YFKKTFP----- 435
DB 1418 RFTPGYTIARRRPLIEDIPEVARALSDSEPADPDGTSALRQTLAGITAPEQHRLLEL 1477
QY 436 ---KAWFKLTH-----RDLGPKS-----RYIGPEVPAEDLIWQDPIP 469
DB 1478 VRSEAATVLTHTRTDDITAGRPFRDGLGDSLTAMELRNRLNTATGLRLPA-TLVFDHPTP 1536
QY 470 AGNTDYCEEVVKQKIAQSGLSISEMVTAWDSARTYRGSDMR--GGANGARIRLAPONEW 527
DB 1537 ORLAGHLHEKLFDSATETALPV---LRTDDDDPIVIVGMACRFPGVVRG-----PEDLM 1587
QY 528 QGNEPERLAKVLSYVEQISADTGASIAADVIVLAGSVGIEKAAKAAGYDVRVPFLKRGDA 587
DB 1588 -----RLLVDSRDEMTDFPADRGWHLAMNA-----FIEESGGA 1621
QY 588 -----TAEMTDADS-----FAPLEPLADGFRNWKQKEYVVRPEEML-----LDRAQL- 629
DB 1622 RQGAFLAEAGDFDAAFPGISPREALA-----MDPOQRLLLETSWELERAGFD 1669
QY 630 -MGLTGPEMTVLLGG 643
DB 1670 PVGLRGSRGTGVFVGG 1684

RESULT 12
US-09-793-708-1
; Sequence 1, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.

APPLICANT: MCDANIEL, Robert
APPLICANT: TANG, Li
TITLE OF INVENTION: RECOMBINANT NARSONOLIDE POLYKETIDE SYNTHASE
FILE REFERENCE: 300622002121
CURRENT APPLICATION NUMBER: US/09/793,708
CURRENT FILING DATE: 2001-02-22
PRIOR APPLICATION NUMBER: US 09/557,440
PRIOR FILING DATE: 2000-09-07
PRIOR APPLICATION NUMBER: US 09/320,878
PRIOR FILING DATE: 1999-05-27
PRIOR APPLICATION NUMBER: US 09/141,908
PRIOR FILING DATE: 1998-08-28
PRIOR APPLICATION NUMBER: US 09/073,538
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 08/846,247
PRIOR FILING DATE: 1997-04-30
PRIOR APPLICATION NUMBER: US 60/134,990
PRIOR FILING DATE: 1999-05-20
NUMBER OF SEQ ID NOS: 38
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 4551
TYPE: PRT
ORGANISM: Streptomyces venezuelae
US-09-793-708-1

Query Match 2.8%; Score 115; DB 11; Length 4551;
Best Local Similarity 21.0%; Pred. No. 2.6;
Matches 138; Conservative 82; Mismatches 232; Indels 204; Gaps 35;
QY 226 APSDERYGVNKPETMENPLAAVQMGLY-----VNPEGVNGH----- 263
Db 624 APTLERV-DVQPVTF-----AVVSLARVWQHGHVTPQAVVGHSGHGETAAAYVAGALS 677
QY 264 PDLR-----TAQVLETF-----RMANDEKTAALTAGG 294
Db 678 DDAARVVTLSKSIHAHLAKGKGLSLALSEDVLERLAGFDGLSVAANGPTATVVS 737
QY 295 -----HTVNCNCHNGNASALADPKASD-----VENQ-----GLGWNPNMOGKASNA 337
Db 738 PVOIELARACADGVRRVIVPDVYASHRSRQVEIIESELAELAGLSPOAPRV-----P 791
QY 338 VTSIGEGATWNTPTFDGMYFOLLFGYNW--ELKKSPPAGAHWEPIDIKKENKPVDS 395
Db 792 FSTLEGAWITEPV-LDGGY-----WYRNLHRVGFAPAVETL-----ATDE 832
QY 396 SIRH-----NPIMTDADMAIKVNPYRAICEKFMADPEYFKKTFKAW---FKLTHRD 446
Db 833 GFTHFEVSAHPVLT---MALPGTVGLATLRDNGGQDLRLVASLAEWANGLAVDWSPL 889
QY 447 GPKSRYIGPEVPA-----EDLIWODPI-----PAGNTDYCEEVVKQIAQ-SGLSISEMVST 497
Db 890 LPSATGHHSDLPYAFQTERHMLGEIEALAPAGEPAVQPAVLRTAAEPAELDRDQLRV 949
QY 498 ANDSA-----RTYR-----GSDMRGANGA-RIRLAPQNEWOG 529
Db 950 ILDKVRAQTAQVLYGATGGQIEVDRTFREAGCTSLTGVDLNRINAAFGVRMAPSMIFDF 1009
QY 530 NEPERLAK--VLSVYEQISADTGASIAADVIVLAGSVGIEKAAAGYDVRVP----- 579
Db 1010 PPEALAEQILLVHVGEAAANPAGAPAPVAAAGAV--DEPVAIVGMACRLPGGVASPED 1067
QY 580 ----FLKGRGDATAEMT-----DADSFAPLEPLADGFRNMOKKEYV-----VKP 619
Db 1068 LWRLVAGGDAISEPPQDRGWDVEGLYHPDPEHPGTSYVRQGGFIENVAGFDAFFGISP 1127
QY 620 EEML-LDRAQLMGLTGPENTVLLGGMRVLGTNYGGTKHGVFTDCGQLTNDFFVNLTD 678
Db 1128 REALAMDPOORLLLETSWEAIVEDAG--IDPISLRGRQGVFT---GAMTHEYGPSLRDGG 1182
QY 679 NSMKPVGSNAYEIRDR---KGAVKWT-----ASRDLVFGSNLSLSYAEVYA 724
Db 1183 E-----GLDGYLLTGNTASVMGRVSYTLGLEGPALTVDYTAC-SSSLVALHLAVQA 1232

RESULT 13
US-10-201-365-2
Sequence 2, Application US/10201365
Publication No. US20030148469A1
GENERAL INFORMATION:
APPLICANT: ASHLEY, Gary
APPLICANT: BETLACH, Melanie C.
APPLICANT: BETLACH, Mary
APPLICANT: MCDANIEL, Robert
APPLICANT: TANG, Li
TITLE OF INVENTION: COMBINATORIAL POLYKETIDE LIBRARIES PRODUCED USING A MODULAR
FILE REFERENCE: PK5 GENE CLUSTER AS SCAFFOLD
CURRENT APPLICATION NUMBER: US/10/201,365
CURRENT FILING DATE: 2002-07-22
PRIOR APPLICATION NUMBER: US 09/141,908
PRIOR FILING DATE: 1998-08-28
PRIOR APPLICATION NUMBER: US 09/073,538
PRIOR FILING DATE: 1998-05-06
NUMBER OF SEQ ID NOS: 32
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 4551
TYPE: PRT
ORGANISM: Streptomyces venezuelae
US-10-201-365-2

Query Match 2.8%; Score 115; DB 12; Length 4551;
Best Local Similarity 21.0%; Pred. No. 2.6;
Matches 138; Conservative 82; Mismatches 232; Indels 204; Gaps 35;
QY 226 APSDERYGVNKPETMENPLAAVQMGLY-----VNPEGVNGH----- 263
Db 624 APTLERV-DVQPVTF-----AVVSLARVWQHGHVTPQAVVGHSGHGETAAAYVAGALS 677
QY 264 PDLR-----TAQVLETF-----RMANDEKTAALTAGG 294
Db 678 DDAARVVTLSKSIHAHLAKGKGLSLALSEDVLERLAGFDGLSVAANGPTATVVS 737
QY 295 -----HTVNCNCHNGNASALADPKASD-----VENQ-----GLGWNPNMOGKASNA 337
Db 738 PVOIELARACADGVRRVIVPDVYASHRSRQVEIIESELAELAGLSPOAPRV-----P 791
QY 338 VTSIGEGATWNTPTFDGMYFOLLFGYNW--ELKKSPPAGAHWEPIDIKKENKPVDS 395
Db 792 FSTLEGAWITEPV-LDGGY-----WYRNLHRVGFAPAVETL-----ATDE 832
QY 396 SIRH-----NPIMTDADMAIKVNPYRAICEKFMADPEYFKKTFKAW---FKLTHRD 446
Db 833 GFTHFEVSAHPVLT---MALPGTVGLATLRDNGGQDLRLVASLAEWANGLAVDWSPL 889
QY 447 GPKSRYIGPEVPA-----EDLIWODPI-----PAGNTDYCEEVVKQIAQ-SGLSISEMVST 497
Db 890 LPSATGHHSDLPYAFQTERHMLGEIEALAPAGEPAVQPAVLRTAAEPAELDRDQLRV 949
QY 498 ANDSA-----RTYR-----GSDMRGANGA-RIRLAPQNEWOG 529
Db 950 ILDKVRAQTAQVLYGATGGQIEVDRTFREAGCTSLTGVDLNRINAAFGVRMAPSMIFDF 1009
QY 530 NEPERLAK--VLSVYEQISADTGASIAADVIVLAGSVGIEKAAAGYDVRVP----- 579
Db 1010 PPEALAEQILLVHVGEAAANPAGAPAPVAAAGAV--DEPVAIVGMACRLPGGVASPED 1067
QY 580 ----FLKGRGDATAEMT-----DADSFAPLEPLADGFRNMOKKEYV-----VKP 619
Db 1068 LWRLVAGGDAISEPPQDRGWDVEGLYHPDPEHPGTSYVRQGGFIENVAGFDAFFGISP 1127
QY 620 EEML-LDRAQLMGLTGPENTVLLGGMRVLGTNYGGTKHGVFTDCGQLTNDFFVNLTD 678
Db 1128 REALAMDPOORLLLETSWEAIVEDAG--IDPISLRGRQGVFT---GAMTHEYGPSLRDGG 1182

QY 679 NSWKPVGSNAYEIRD-----KTGAVKWT-----ASRVDLVFGSNLSRSYAERYA 724
Db 1183 E-----GLDGYLLTGNTASVMSGRVSYTLGLEGPALTVDTAC-SSSLVALHIAVOA 1232

RESULT 14
US-10-160-539-1
; Sequence 1, Application US/10160539
; Publication No. US20030162262A1
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062200210
; CURRENT APPLICATION NUMBER: US/10/160,539
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: US/09/657,440
; PRIOR FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908
; PRIOR FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 4551
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-10-160-539-1

Query Match 2.8%; Score 115; DB 12; Length 4551;
Best Local Similarity 21.0%; Pred. No. 2.6;
Matches 138; Conservative 82; Mismatches 232; Indels 204; Gaps 35;

QY 226 APSDERYGVNKPETMENPLAAVOMGLY-----VNPEGVNGH----- 263
Db 624 APTLERV-DVVQPVTF-----AVMVSARVWQHGVTPQAVVGHSGQSEIAAAYVAGALS 677

QY 264 PDPLR-----TAQQVLETF-----RMANDEKTAALTAGG 294
Db 678 DDAARVTVLRKSTAAHLAGGMLSLALSDAVLERLAGDGLSVAANGPTATVVS 737

QY 295 -----HTVGNCHGNASALAPDPKASD-----VENO-----GLGNGNPNMOGKASNA 337
Db 738 PVQTEELARACEADGVRAVRIPVDYASHRSQVEIIESELAELVAGLSPOAPRV-----P 791

QY 338 VTSIGEGAWTNTPTKFDGMYFDLLFGYNW--ELKKSPAGAHWEPIDIKKENKPVDSADP 395
Db 792 FFSTLEGAWITEPV-LDGGY-----WYRNLRHRVGFAPAVETL-----ATDE 832

QY 396 SIRH-----NPIMTDADMAIKVNPYRAICEKPMADPEYFKKTFKAW-----FKLTHRD 446
Db 833 GFTHFEVSAHPVLT---MALPGVTGLATLRDNGGQDRLVLAELAEAWGLAVDSW 889

QY 447 GPKSRYIGPEVPA-----EDLIWDP-----PAGNTDYCEEVVKOKIAQ-SGLSISEMVS 497
Db 890 LPSATGHSDLPVAFQTERHWLGEIEALAPAGEPAVQPAVLRTAEAPAEALDRDEQLR 949

QY 498 AWDSA-----RTYR-----GSDMRGGANGA-RIRLAPQNEWOG 529
Db 950 ILDKVRAQTAQVLGYATGGQIEVDRTFREAGCTSLTGVDLRNRINAAFGVMAFSPMIFDF 1009

QY 530 NEPERLAK--VLSVYEQISADTGASIAADVIVLAGSVGIERAKAAGYDVRVP----- 579
Db 1010 PTEPALAEQLLVVHGHAANPAGAEAPVAAAGAV--DEPVAIVMACRLPGGVASPED 1067

QY 580 ----FLKGRGDATAMT-----DADSFAPLEPLADGFRNWKQKEV-----VKP 619
Db 1068 LWRLVAGGGDAISEFFQDRGDWDEGLYHPDPEHPTGTSYVROGGFIENVAGFAAFAFGISP 1127

QY 620 EEML-LDRAQLMGLTGPEMTVLLGGMRVLTGYGTTKGHGVFTDCEGQLTNDFFVNLDMG 678
Db 1128 REALANDPOORLLLETSWAEVDAG--IDPTSLRGRGVGVFT---GAMTHEYGPSLRDGG 1182

QY 679 NSWKPVGSNAYEIRD-----KTGAVKWT-----ASRVDLVFGSNLSRSYAERYA 724
Db 1183 E-----GLDGYLLTGNTASVMSGRVSYTLGLEGPALTVDTAC-SSSLVALHIAVOA 1232

RESULT 15
US-09-861-289-31
; Sequence 31, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 4613
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-861-289-31

Query Match 2.8%; Score 115; DB 10; Length 4613;
Best Local Similarity 21.0%; Pred. No. 2.6;
Matches 138; Conservative 82; Mismatches 232; Indels 204; Gaps 35;

QY 226 APSDERYGVNKPETMENPLAAVOMGLY-----VNPEGVNGH----- 263
Db 696 APTLERV-DVVQPVTF-----AVMVSARVWQHGVTPQAVVGHSGQSEIAAAYVAGALS 739

QY 264 PDPLR-----TAQQVLETF-----RMANDEKTAALTAGG 294
Db 740 DDAARVTVLRKSTAAHLAGGMLSLALSDAVLERLAGDGLSVAANGPTATVVS 799

QY 295 -----HTVGNCHGNASALAPDPKASD-----VENO-----GLGNGNPNMOGKASNA 337
Db 800 PVQTEELARACEADGVRAVRIPVDYASHRSQVEIIESELAELVAGLSPOAPRV-----P 853

QY 338 VTSIGEGAWTNTPTKFDGMYFDLLFGYNW--ELKKSPAGAHWEPIDIKKENKPVDSADP 395
Db 854 FFSTLEGAWITEPV-LDGGY-----WYRNLRHRVGFAPAVETL-----ATDE 894

QY 396 SIRH-----NPIMTDADMAIKVNPYRAICEKPMADPEYFKKTFKAW-----FKLTHRD 446
Db 895 GFTHFEVSAHPVLT---MALPGVTGLATLRDNGGQDRLVLAELAEAWGLAVDSW 951

QY 447 GPKSRYIGPEVPA-----EDLIWDP-----PAGNTDYCEEVVKOKIAQ-SGLSISEMVS 497
Db 952 LPSATGHSDLPVAFQTERHWLGEIEALAPAGEPAVQPAVLRTAEAPAEALDRDEQLR 1011

QY 498 AWDSA-----RTYR-----GSDMRGGANGA-RIRLAPQNEWOG 529
Db 1012 ILDKVRAQTAQVLGYATGGQIEVDRTFREAGCTSLTGVDLRNRINAAFGVMAFSPMIFDF 1071

QY 530 NEPERLAK--VLSVYEQISADTGASIAADVIVLAGSVGIERAKAAGYDVRVP----- 579
Db 1072 PTEPALAEQLLVVHGHAANPAGAEAPVAAAGAV--DEPVAIVMACRLPGGVASPED 1129

QY 580 ----FLKGRGDATAMT-----DADSFAPLEPLADGFRNWKQKEV-----VKP 619
Db 1130 LWRLVAGGGDAISEFFQDRGDWDEGLYHPDPEHPTGTSYVROGGFIENVAGFAAFAFGISP 1189

QY 620 EEML-LDRAQLMGLTGPEMTVLLGGMRVLTGYGTTKGHGVFTDCEGQLTNDFFVNLDMG 678

Db 1190 REALAMPQORLLLETWEAVEDAG--IDPTSLRGQGVFT---GANTHEYGFSLRDGG 1244
Qy 679 NSWKPVGSGNAYEIRDR----KTGAVKWT-----ASRVDLVFGSNSLLRSYAEVYA 724
Db 1245 E-----GLDGYLLTGNTASVMSGRVSYTLGLEGPALTVDTAC-SSSLVALHLAVQA 1294

Search completed: October 7, 2003, 19:59:24
Job time : 52.759 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 7, 2003, 19:39:02 ; Search time 29.6575 Seconds
(without alignments)
2441.707 Million cell updates/sec

Title: US-09-884-889-6
Perfect score: 4058
Sequence: 1 MNNASADDLHSSLQRCRAF.....RDFVAATWTKVMNADRFVAS 753

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2919	71.9	724	2 B82186	catalase/peroxidase
2	2820	69.5	723	2 AF3127	catalase [imported]
3	2820	69.5	731	2 C98160	catalase (AB033631)
4	2795	68.9	754	2 S75113	catalase (BC 1.11.)
5	2731	67.3	720	2 S71130	catalase (BC 1.11.)
6	2514	62.0	735	2 B83763	catalase BH0906 [1
7	2489	61.3	751	1 JS0520	catalase (BC 1.11.)
8	2489	61.3	751	4 T43873	catalase (BC 1.11.)
9	2489	61.3	751	4 T43874	catalase (BC 1.11.)
10	2489	61.3	751	4 T43878	catalase (BC 1.11.)
11	2489	61.3	751	4 T43880	catalase (BC 1.11.)
12	2489	61.3	751	4 T43881	catalase (BC 1.11.)
13	2489	61.3	752	4 T43877	catalase (BC 1.11.)
14	2489	61.3	753	4 T43882	catalase (BC 1.11.)
15	2279	56.2	781	2 F82584	catalase/peroxidase
16	2255	55.6	727	2 F95422	catalase (BC 1.11.)
17	2207	54.4	740	2 T45091	catalase (BC 1.11.)
18	2162.5	53.3	737	2 A78626	catalase/peroxidase
19	2096	51.7	726	2 G91237	hydroperoxidase HP
20	2096	51.7	726	2 C86085	catalase, hydropor
21	2095	51.6	726	1 CSECHP	catalase (BC 1.11.)
22	2083	51.3	726	2 A10936	catalase (hydropor
23	2063.5	50.9	746	1 A43685	catalase (BC 1.11.)
24	2051.5	50.6	727	1 CSEBHT	catalase (BC 1.11.)
25	2042.5	50.3	740	1 A40662	catalase (BC 1.11.)
26	1998	49.2	731	2 T44846	catalase (BC 1.11.)
27	1930	47.6	720	2 T44562	catalase (BC 1.11.)
28	1922.5	47.4	737	2 AC0403	catalase (BC 1.11.)
29	1901.5	46.9	741	2 A69529	catalase (BC 1.11.)

ALIGNMENTS

RESULT 1

B82186
catalase/peroxidase VC1560 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 24-Aug-2001
C:Accession: B82186
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: B82186
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-724 <HEI>
A:Cross-references: GB:AE004233; GB:AE003852; MID:g9656055; PIDN:AAF94714.1; GSPDB:GN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC1560
A:Map position: 1
C:Superfamily: catalase HPI
C:Keywords: heme; iron; metalloprotein
F;99/Active site: His (distal axial ligand) #status predicted
F;267/Binding site: heme iron (His) (proximal axial ligand) #status predicted
F;318/Active site: Trp (tryptophyl radical intermediate) #status predicted
Query Match 71.9%; Score 2919; DB 2; Length 724;
Best Local Similarity 74.1%; Pred. No. 5.2e-194;
Matches 530; Conservative 75; Mismatches 110; Indels 0; Gaps 0;
QY 37 SGKCPVHGNTSTGTSTKNDWPEGLNLDLTHOODRSDPMDPDFNVREVRKLDLALK 96
DB 9 SGQCPVHGGLTSASNSMDWMPKALNLDLTHODSKTNPLGADFNYREELKLDVALK 68
QY 97 KDVHALMTSQEWPPADWGHYGGIMTMAHWSAGTYRIADRGGGTGSORFAPLNSWPD 156
DB 69 RDLKALMTNSQEWPPADWGHYGGIMTMAHWSAGTYRIADRGGGTGNORFAPLNSWPD 128
QY 157 NVSLDKARLLWPIKKYGNKISWADLMLIAGTVAYESMGLPAYGFSFGRVDIWEPEKDI 216
DB 129 NANLDKARLLWPIKKYGNKISWADLMLIAGTVAYESMGLPAYGFSFGRVDIWEPEKDI 188
QY 217 YGDEKEWLAPSDERYGVNKPETMENPLAAVOMGLIYVNPGEVNGHPDPLRTAQVLET 276
DB 189 YGSEKEWLAKSGGNSRYSQGRDLENPLAAVOMGLIYVNPGEVNGHPDPLRTAQVLET 248
QY 277 FARMANDEKTAALTAGGTVGCHNGNASALAPDPKASDVENQGLGWGNPNMOGRASN 336
DB 249 FARMANDEKTAALTAGGTVGCHNGNASALAPDPKASDVENQGLGWGNPNMOGRASN 308
QY 337 AVTSGIEGAWTTNPTKFDWGMGYFDLLFCYNWELKKSPAGAHHWEPIDIIKENKPVDSAPS 396

```
Db 309 TTVSGIEGAWTHPTRWNEFYLLSYEWQLTKSPAGAWQWEPVNIKEEDKPDVDEPS 368
Qy 397 IRHNPIMTDADMAIKVNPYRAICBKFADPEYFKKTKFAKAWFKLTHRDLPKPSRYIGPE 456
Db 369 IRYNPMWTDADMAIKIDPEYKRISEYFKDPAFSEYFARAWFKLTHRDMPKARYFGPD 428
Qy 457 VPAEDLIWQDPIPAAGTNDYCEVYKQKIAQSGLSISEMVSMTAWSDARTYRSGDMRGANG 516
Db 429 VPAEDLIWQDVPAGRKDYDYNVAKAKIAAAGLSISEMVSMTAWSDARTFRGSDKRGANG 488
Qy 517 ARIRLAPQNEQWQNEPERLAKVLSYEQISADTASIAADVIVLAGSVGIEKAAKAGYDV 576
Db 489 ARIRLAPQWQNEPARGKVLAVLEKIAESGISIADTIVLAGNVGIEQAAKAGYDV 548
Qy 577 RVPLKRGDGAETADSEAPLEADGFRNMQKKEVYVVKPEMLLDRAQLMGLTGE 636
Db 549 TVPAPRGDGAETADSEAPLEADGFRNMQKKEVYVVKPEMLLDRAQLMGLTGE 608
Qy 637 MTVLGGMRLVGTNYGKKGHTDCBQGLNDFVNLTDGMSKPKVGSNAYEIRDRKT 696
Db 609 MTVLGGMRLVGTNYGKKGHTDCBQGLNDFVNLTDGMSKPKVGSNAYEIRDRKT 668
Qy 697 GAVKWTASRDVLFGSNLSRSYAEVYAADNGEKFVDFVAAWTKVMNAADRFV 751
Db 669 GKVRWATRDVLFGSNLSRSYAEVYAADNGEKFVDFVAAWTKVMNAADRFV 723

RESULT 2
AF3127
catalase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AF3127
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan,
Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:21608550; PMID:11743193
A:Accession: AF3127
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-723 <KUR>
A:Cross-references: GB:AF008689; PIDN:AAL45436.1; PID:g17743138; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: katA
A:Map position: linear chromosome
C:Superfamily: catalase HPI

Query Match 69.5%; Score 2820; DB 2; Length 723;
Best Local Similarity 71.0%; Pred. No. 3.7e-187;
Matches 508; Conservative 86; Mismatches 121; Indels 0; Gaps 0;

Qy 37 SGKCPVMHGGTSTGTSNKNWPEGLNLDILHQODRKSDDPMDPDFNYEEVYKLDLDFALK 96
Db 8 AGKCPVMHGGTASGSKSVTEWPNALNLDILHQHDTKTNPLGTSFYREALKTLDEALK 67
Qy 97 KDVALMTDSQEWNPADMGHYGGLMIRMAWISAGTYRTADGGGGTGSQRAPLNSWPD 156
Db 68 ADRLALMTDSQEWNPADMGHYGGLMIRMAWISAGTYRTADGGGGTGSQRAPLNSWPD 127
Qy 157 NVSLDKARLLWPKIKKYGKISWADLMLAGTVAYESMGLPAYGFSFGVDIWEPEKDI 216
Db 128 NVNTDKGRLLWPKIKKYGKISWADLMLAGTVAYESMGLPAYGFSFGVDIWEPEKDI 187
Qy 217 YWDEKEWLAPSDERYGDVKNKPEWNPENLAQVOMGLIYVNPBGVNGHPDPLRTAQOVL 276
Db 188 YWDEKEWLAPSDERYGDVKNKPEWNPENLAQVOMGLIYVNPBGVNGHPDPLRTAQOVL 247
```

```
Qy 277 FARMANDEKTAALTAGTHTVGNCHGNASALAPDPKASDVENQGLGWNHNPQMGKASN 336
Db 248 FARMANDEKTAALTAGTHTVGNCHGNASALAPDPKASDVENQGLGWNHNPQMGKASN 307
Qy 337 AVTSGIEGAWTNTKPKFDMGYFDLLFGYNWELKSPAGAHWEPIIDIKENKPDVADSPS 396
Db 308 TVVSGIEGAWTSEPTKWDNGFFDMLFKHEWTLTHSPAGASQWAPITIAEEDKPDVDEAS 367
Qy 397 IRHNPIMTDADMAIKVNPYRAICBKFADPEYFKKTKFAKAWFKLTHRDLPKPSRYIGPE 456
Db 368 IRYNPMWTDADMAIKIDPEYKRISEYFKDPAFSEYFARAWFKLTHRDMPKARYFGPD 427
Qy 457 VPAEDLIWQDPIPAAGTNDYCEVYKQKIAQSGLSISEMVSMTAWSDARTYRSGDMRGANG 516
Db 428 VPAEDLIWQDPIPAAGTNDYCEVYKQKIAQSGLSISEMVSMTAWSDARTFRGSDKRGANG 487
Qy 517 ARIRLAPQNEQWQNEPERLAKVLSYEQISADTASIAADVIVLAGSVGIEKAAKAGYDV 576
Db 489 ARIRLAPQWQNEPARGKVLAVLEKIAESGISIADTIVLAGNVGIEQAAKAGYDV 547
Qy 577 RVPLKRGDGAETADSEAPLEADGFRNMQKKEVYVVKPEMLLDRAQLMGLTGE 636
Db 549 TVPAPRGDGAETADSEAPLEADGFRNMQKKEVYVVKPEMLLDRAQLMGLTGE 607
Qy 637 MTVLGGMRLVGTNYGKKGHTDCBQGLNDFVNLTDGMSKPKVGSNAYEIRDRKT 696
Db 609 MTVLGGMRLVGTNYGKKGHTDCBQGLNDFVNLTDGMSKPKVGSNAYEIRDRKT 667
Qy 697 GAVKWTASRDVLFGSNLSRSYAEVYAADNGEKFVDFVAAWTKVMNAADRFV 751
Db 669 GKVRWATRDVLFGSNLSRSYAEVYAADNGEKFVDFVAAWTKVMNAADRFV 722

RESULT 3
C98160
catalase (AB033631) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C:Accession: C98160
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorillo, B.; Goldm
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz,
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; PMID:21608551; PMID:11743194
A:Accession: C98160
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-731 <KUR>
A:Cross-references: GB:AE007870; PIDN:NAK88805.1; PID:g15158560; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L481
A:Map position: linear chromosome
C:Superfamily: catalase HPI

Query Match 69.5%; Score 2820; DB 2; Length 731;
Best Local Similarity 71.0%; Pred. No. 3.7e-187;
Matches 508; Conservative 86; Mismatches 121; Indels 0; Gaps 0;

Qy 37 SGKCPVMHGGTSTGTSNKNWPEGLNLDILHQODRKSDDPMDPDFNYEEVYKLDLDFALK 96
Db 16 AGKCPVMHGGTASGSKSVTEWPNALNLDILHQHDTKTNPLGTSFYREALKTLDEALK 75
Qy 97 KDVALMTDSQEWNPADMGHYGGLMIRMAWISAGTYRTADGGGGTGSQRAPLNSWPD 156
Db 76 ADRLALMTDSQEWNPADMGHYGGLMIRMAWISAGTYRTADGGGGTGSQRAPLNSWPD 135
Qy 157 NVSLDKARLLWPKIKKYGKISWADLMLAGTVAYESMGLPAYGFSFGVDIWEPEKDI 216
Db 136 NVNTDKGRLLWPKIKKYGKISWADLMLAGTVAYESMGLPAYGFSFGVDIWEPEKDI 195
Qy 217 YWDEKEWLAPSDERYGDVKNKPEWNPENLAQVOMGLIYVNPBGVNGHPDPLRTAQOVL 276
Db 196 YWDEKEWLAPSDERYGDVKNKPEWNPENLAQVOMGLIYVNPBGVNGHPDPLRTAQOVL 255
```


142	QY	GTGSRQFAPLNSWPONVSLDKARLLNPIKKKYGNKISWADMLILAGTVAYESMGILPAYG	201
138	Db	ATGNORFAPLNSWPONVNLDKARLLNPIKKKYGNKLSWGLDILITAGTMAYESMGILKVYG	197
202	QY	FSFGVRDITWEPKDIYWGDEKWLAPSDERYGVNKPETMENPLAAVOMGLIIVNPEGVN	261
198	Db	FAGGREDIWHPEKDIYWGAEKWLASSDHRITGSSDR--ESLENPLAAVOMGLIIVNPEGVD	256
262	QY	GHPDPLRTAQQLYETFARMAMNDEKTAALTAGGHTVNGCHGNASALAPDPKASVQNG	321
257	Db	GHPDPLCTAQDVRTTFARMAMNDEETVALTAGGHTVGKCHGNSKAELIGPEPEGADVWEQ	316
322	QY	GLWGNGPNWOGKASNAVTSGIEGATWTPNPFDMGYEDLLFEGYNWELKSPACAGAHWEPI	381
317	Db	GLWGNAQNGKVGRTSSGIEGATWTHPTQMDNGSYVTLFNHEWELKSPACAGAWQEPV	376
382	QY	DIKKENKPVDA5DPDSIRINPIMTDADMAIKVNPYTRAICEKFMADPEYFKTKAKAWFL	441
377	Db	NIKEEDKPVQVEDPNIRNPIMTDADMAIKDPIYRQISRRFYREPDYFAEVFAKAWFL	436
442	QY	THRDILGPKSRYIGPEVPAEDLIWQDPIPNAGWNTDYCEBWK---OKIAQSGLSISEMVSTA	498
437	Db	THRDILGPKSRYIGPVPQVEDLIWQDPIPNVDTYITSEGEIKELEQQIILASGLTVSELVCTA	496
499	QY	WDSARTYRGSDMRGANGARIRLAPONEWOGNEPERLAKVLSVYVEQISADTG--ASTADY	556
497	Db	WDSARTFRSSDYRGANGARIRLEPQKNWQNEPTRLAKVLAVLVENTQANFAKPVSIADL	556
557	QY	IYLAGSVGIEKAAGAAGYDVRVPFLKRGDATAEMTDADSFAPLEPLADGFRNQKKEYV	616
557	Db	IYVGGGAAATAALDGGIEVNVVFLPGRGDATAQMTDAESFTPLPETHDRNWLKODYA	616

```

617 V A P E E M L L D R A U L A M G L T G P E M T V L D G G M R V L G T A Y G G T K R G V F P D C E G Q T I N D F F V A L I T D 87
617 V S P E E L L L E R T Q L M G L T A P E M T V L I G G M R V L G T N H G G T K H G V F T D R V G V L S N D F F V N L T D 676
677 M G N S K P V G S N A Y E I R D A K T G A V K T A S R V D L V F G S N L L S R S Y A E V T A Q D D N G E K F V R D F 736
677 M A Y Q H R P A G N N I Y E I G D R Q T G E V K W T A T K V D L V F G S N L L S R S Y A E V T A Q D D N R E K F V R D F 736
737 V A A W T K V M N A D R E D V 751
737 V A A W T K V M N A D R E D L 751

RESULT 5
S71130
catalase (EC 1.11.1.6) - Synecchococcus sp. (strain PCC 7942)
N:Alternate names: catalase-peroxidase
N:Contains: peroxidase (EC 1.11.1.7)
C:Species: Synecchococcus sp.
A:Variety: PCC 7942
C:Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000
C:Accession: S71130; S77717
R:Mutisuda, M.; Ishikawa, T.; Takeda, T.; Shigeoka, S.
Biochem. J. 316, 251-257, 1996
A:Title: The catalase-peroxidase of Synecchococcus PCC 7942: purification, nucle
A:Reference number: S71130; MUID:96235199; PMID:8645214
A:Accession: S71130
A:Molecule type: DNA

```

A; Residues: 1-720 <MULT>
A; Cross-references: EMBL: D61378; NID: gi550616; PID: gi1359464
A; Experimental source: PCC 7942
A; Accession: S77717
A; Molecule type: protein
A; Residues: 2-7; 9-15 <MULT>
C; Comment: This catalase is not sensitive to 3-amino-1,2,4-triazole.
C; Superfamily: catalase HPI
C; Keywords: chromoprotein; heme; homodimer; iron; metalloprotein; oxidoreducta
F; 2-720/Product: catalase #status experimental <NAT>
F; 94/Active site: His (distal axial ligand) #status predicted
F; 262/Binding site: heme iron (His) (proximal axial ligand) #status predicted

F:313/Active site: Trp (tryptophyl radical intermediate) #status predicted

Query Match 67.3%; Score 2731; DB 2; Length 720;
Best Local Similarity 69.8%; Pred. No. 5.3e-181;
Matches 504; Conservative 83; Mismatches 125; Indels 10; Gaps 5;

[illegible]

RESULT 6
B83763
catalase BH0906 (imported) - *Bacillus halodurans* (strain C-125)

C;Species: *Bacillus narodurans*
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: B83763

R; Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirata, T. *Nucleic Acids Res.* 28, 4317-4331, 2000

A: Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and A: Reference number: A83650; MUID:20512582; PMID:11058132

A: Accession: B83763

A: Status: preliminary

A: Molecule type: DNA

A: Residues: 1-735 <STO>

A; Cross-references: GB:AP001510; GB:BA000004; NID:G10173440; PIDN:BA04625.1; GSPDB:G
A; Experimental source: strain C-125

C;Genetics:

A:Gene: BH0906

C: Superfamily: catalase HPT

Query Match 62.0%; Score 2514; DB 2; Length 735;
Best Local Similarity 64.1%; Pred. No. 5.8e-166;
Matches 464; Conservative 90; Mismatches 156; Indels 1

QY	37	SGKCPVHHGNGT---STGTSNKDWWPEGLNDILHQODRKSDPMDPDFNYREEVRKLD93
Db	6	TKGCPFHGATTSKSSSGTNNDDWPNALNILRQHDKKSNPMBEEDYAEFEESKLDYD 65
QY	94	ALKKDYHALMTDSQEWADRGHVGGLMIRMAWHSAGTYRIADRGGGGTSORFAPLNS 153
Db	66	ALKQDVRDLMRDSQDWWFADTGHYGPFFIRMSWHAAGTYRIGDRGGGGTGNORFAPLNS 125
QY	154	WPDNVSLDKARLLWPIPKKYGKNSWADMLILAGTVAYESMGLPAYGFGSEGRVDLWEPE 213
Db	126	WPDNGNLDKARLLWPIKQYGNKISWADLLVYLAGNVAIEDMGGPVIGFCAGREDLWHPPE 185
QY	214	KDIYWGDEKEWLAPSDERY-GDVNKPETMENPLAAVQMLTYVNPESVNGHPDPLRTAQO 272
Db	186	EDIYWGSEKWLTDGKRYSGD---RELENPLAAVEWGLIYVNPPEGDPCKPIKAAHD 240
QY	273	VLETFARMANDEKTAALTAGHTVGNCHNGNNASALAPDKKASDVENQGLGHCNPNMOG 332
Db	241	IRETFRGMNDDEETVALIAGHTFGKAGAGNPDHVGPEPEAPTEAOGSLGQNTYGS 300
QY	333	KASNAVTSIEGAVTNTPTKFDMGYFDLLFGYVWELKKSPAGAAHWEPIDIKKENKPVDA 392
Db	301	KGRDVTISLEGAWTANPTQDNGCFDILLFCYEWWLTKSPAGAYQWQAVDPDEKDLAPDA 360
QY	393	SPDSIRHNPIWTDADMAIKVNPTYRACEKPMADPEYFKTKFKAAMFKLTHRDLGPKSRY 452
Db	361	EDPSKVPVTMTDLALRHDPEYEKISRREHKHKNDFADAFARAWFKLLHRDMGPKARY 420
QY	453	IGPEVPAEDLIWQDIPAGN---TDYCEEVVKQIAOSGLSISEMYSTANDSARTYRGSD 509
Db	421	LGPEVPAEDFIWQDVPVTVDLDAEVEELKAKILDSGLTVSELVTTAWASASTFRNSD 480
QY	510	MRRGGANGARILAPQNEHQGNEPERLARLSVYEQISA--DTGASITADVLVLAGSVGIBK 567
Db	481	KRGGANGARILAPQKDWNEVQPOLEKVLSVLENIOSQLDKKVSITADLVLGGSAAVEK 540
QY	568	AKAKAGYDVRVDFLKGGRDATAEMTDAOSFAPLEPLADGFRNMOKKEYVYVVKPEMLLDRA 627
Db	541	AAKEAGFDVTFPAPGRGDATOEOTDEBGFVLEPVSDGFRNYOKKEYSVSPPELLIDKA 600
QY	628	QLMGLTGPEMTVLLGGMVLTNTYGGYKHGVFTDCEQLTNDNFVNLTDMGNSWKPKVGSN 687
Db	601	QLLDUTAPMTALJGGMALCANTGGYQHGVFTDCVGTLTNDFVNLDDMGIEWKPVVDYN 660
QY	688	AYEIRDRKTGAVKMTASRVDLVFGSNSLLRSYAEVYAADONGEKFYRDFVAAMTKVYNNAD 747
Db	661	LYEGDRKTEGVVRTATRVDLVFGSNSITLRAIEVYAADONKCKGFVEDFTAAWVKVYNNAD 720
QY	748	REFDV 751
Db	721	REFDL 724

RESULT 7
JS0520

catalase (EC 1.11.1.6) HPI - *Bacillus stearothermophilus*
 N:Alternate names: catalase-peroxidase; hydroperoxidase I
 N:Contains: peroxidase (EC 1.11.1.7)
 C:Species: *Bacillus stearothermophilus*
 C:date: 30-Sep-1991 #sequence_revision 08-Nov-1996 #text_
 C:Accession: J05020; A33955
 R:Trakulnalemsai, S.; Aihara, S.; Miyai, K.; Suga, Y.; Y
 submitted to JPIID, July 1991

A:Reference number: JS0520

A:Accession: JS0520

A:Molecule type: DNA

A:Residues: 1-735 <TRA>

R:Loprasert, S.; Negoro, S.; Okada, H.

J. Bacteriol. 171, 4871-4875, 1989

A:Title: Cloning, nucleotide sequence, and expression in *Escherichia coli* of the *Bacillus*

A:Reference number: A33955; MUID:89359121; PMID:2670897

A:Accession: A33955

A:Molecule type: DNA

A:Residues: 1-731 <LOP>

A:Note: this sequence has been extensively revised in reference JS0520

C:Genetics:

A:Gene: perA

C:Complex: homotetramer; two non-covalently associated iron protoheme IX groups per tetra

A:Description: <CAT>

C:Function: as catalase, catalyzes the dismutation of two molecules of hydrogen perox

C:Function: <PER>

A:Description: as peroxidase, uses hydrogen peroxide to oxidize donor compounds produc

C:Superfamily: catalase HPI

C:Keywords: chromoprotein; heme; homotetramer; iron; metalloprotein; oxidoreductase

F:101/Active site: His (distal axial ligand) #status predicted

F:264/Binding site: heme iron (His) (proximal axial ligand) #status predicted

F:315/Active site: Trp (tryptophyl radical intermediate) #status predicted

Query Match

Best Local Similarity 61.3%; Score 2489; DB 1; Length 735;

Matches 468; Conservative 85; Mismatches 160; Indels 16; Gaps 7;

```
QY 32  RERAMSGKCPVHGGNT---STGTSNKDWPEGLNLDLHQDRKSDPMDPDFNYREVR 88
DB 4  QNRQNAACCP-FHGSVTNQSSNRNTNKDWPNQLNLSILHQDRKTNPHDEEFNAYEEFQ 62
QY 89  KLDFDALKDKYHVALMTDSQEWPPADWGHYGLMIRMAHWSAGTYRIADRGSGGTGSGRF 148
DB 63  KLDYWALKEDLRLKLTESQDWWPADYGHYGLPIFIRMAHWSAGTYRIGDRCGASTGTGRF 122
QY 149  APLNSWPNVSLDKARRLLWPIKKYGNKISWADLMILAGTVAYESMGLPAYGFSFRVD 208
DB 123  APLNSWPNANLDKARRLLWPIKKYGNKISWADLFILAGNVAIESMGGKTIIGFGRVD 182
QY 209  IWEPEKDIYGDKEWLAPESDERY-GDVNKPETHENPLAAVOMGLIYVNPGEVNGHPDPL 267
DB 183  VWHPEEDYWGSEKWL---SERYSGD----RELENPLAAVOMGLIYVNPGEVNGHPDPL 236
QY 268  RTAQVLETFARMAMNDEKTAALTAGGHTVGNCHGNASALAPDPKASDVENQGLGWN 327
DB 237  AAARDIRFTFRMGNDDETVALLAGGHTFGKAGAGATHVGPPEAPPEAQLGLWIS 296
QY 328  PNMCKASNAVTSIGEGAWTNTPTKFDMGYFDLLFGYNWELKSPAGAHHPEDIDIKEN 387
DB 297  SYGKKGSDTITSGIEGAWTPTPTQWDTSYFDMLFGYDWLTKSPAGAWQWMAVDPDEK 356
QY 388  KPVASDPSIRHNPTMTDADMAIKVNPYRICEKFMADPEYFKKTFKAKFWKLTDRDLG 447
DB 357  LAPDAEDPSKVPNTMMTDLALRFDPEYKIRRFHONPEEFAFAFAFWFKLTDRDMG 416
QY 448  PKSRVIGVEPAEDLIWODPIAGNTDYCE---EVVKOKIAQSGLSISEMSTANDSART 504
DB 417  PKTRYLGVEPKEDFIWQDPIPEVDYELTEAIEIEIKALINSLGLTVSELVKTAWASAST 476
QY 505  YRSDMRGGANGARIRLAPQWNOGNEPERLAKVLSVYEQISAD--TGASIADEVIVLAGS 562
DB 477  FRNSDKRGANGARIRLAPQKDWNEVNEPERLAKVLSVYEDIORELPPKKVSIADLIVLGS 536
QY 563  VGIEKAAKAAGYDVVPFLKRGDATAEMTADSFAPLEADGFRNWKQKEYVYVVKPEEM 622
DB 537  AAVEKAARDAGDFVKVPPFGRGDQTQDVFESFVLEPFDAGFRNTQKQFYSVYPEEL 596
QY 623  LLDRAQLMGLGPMTVLLGGMVRLGTYNGKGVFTDCRGOLTNDFEVLNLTGMNSWK 682
DB 597  LVDKAQLGLTAPMTVLLGGLRVLGANTRLPHGVFTDRIGVITLNDFFVLLDMNEMV 656
```

```
QY 683  PVGSNAYEIRDRKTVGAVKWTASRYVDLVFGSNLSLSRYAETVAQDDNGEKFYRDFVAAWTK 742
DB 657  PTDSGIYEIRDRKTEVRWTATRYDLIFGNSILRSYAEFYAQDDNQEKYRDFINAWVK 716
QY 743  VMNADREFDV 751
DB 717  VMNADREFDL 725
```

RESULT 8

T43873

catalase (EC 1.11.1.6) [imported] - synthetic

C:Species: synthetic

A:Note: cat gene engineered and expressed in *Bacillus stearothermophilus*

C:Date: 24-Mar-2000 #sequence_revision 24-Mar-2000 #text_change 23-Mar-2001

C:Accession: T43873

R:Matsuura, T.; Miyai, K.; Trakulnaleamsai, S.; Yomo, T.; Shima, Y.; Miki, S.; Yamam

Nat. Biotechnol. 17, 58-61, 1998

A:Title: Evolutionary molecular engineering by random elongation mutagenesis.

A:Reference number: Z22644; MUID:99116998; PMID:9920270

A:Accession: T43873

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-751 <MAT>

A:Cross-references: EMBL:AB020075; NID:g4140057; PIDN:BAA36987.1; PID:g4140058

C:Comment: For the wild-type sequence, see PIR:JS0520.

C:Genetics:

A:Gene: cat

C:Keywords: oxidoreductase

Query Match 61.3%; Score 2489; DB 4; Length 751;

Best Local Similarity 64.2%; Pred. No. 3.2e-164;

Matches 468; Conservative 85; Mismatches 160; Indels 16; Gaps 7;

```
Db 537 AAVEKAARDAGFDVKVPPFFGRCGDATQEQDQDVSFAVLEPFADGFRNYQKQESVPPPEL 596
QY 623 LLDRAQLMGLTGPEMTVLLGGMVLTGNTYGGTGHGVTDCGQLTNDFFVNLNLDGMSWK 682
Db 597 LVDKAQLGLGTAPMTVVLGGLRVLCANYRDLPHGVFTDRIGVLTNDFFVNLNLDGMYEW 656
QY 683 PVGSNAYEIRDRKTGAVKWTASRVDLVFGSLSLLRSYAEEYAQDDNGEKFVRDFVAATK 742
Db 657 PTDSGIYEIRDRKTGEVRWTATRVDLIFGNSILRSYAEEYAQDDNGEKFVRDFINAWK 716
QY 743 VMNADRFV 751
Db 717 VMNADRFDL 725

RESULT 9
T43874
catalase (EC 1.11.1.6) [imported] - synthetic
C:Species: synthetic
A:Note: cat gene engineered and expressed in Bacillus stearothermophilus
C:Date: 24-Mar-2000 #sequence_revision 24-Mar-2000 #text_change 23-Mar-2001
C:Accession: T43874
R:Matsumura, T.; Miyai, K.; Trakulnaleaamsai, S.; Yomo, T.; Shima, Y.; Miki, S.; Yamamoto,
Nat. Biotechnol. 17, 58-61, 1998
A:Title: Evolutionary molecular engineering by random elongation mutagenesis.
A:Reference number: Z22644; MUID:99116998; PMID:9920270
A:Accession: T43874
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-751 <MAT>
A:Cross-references: EMBL:AB020079; NID:94140065; PIDN:BAA36991.1; PID:94140066
C:Comment: For the wild-type sequence, see PIR:JS0520.
C:Genetics:
A:Gene: cat
C:Keywords: oxidoreductase

Query Match 61.3%; Score 2489; DB 4; Length 751;
Best Local Similarity 64.2%; Pred. No. 3.2e-164;
Matches 468; Conservative 85; Mismatches 160; Indels 16; Gaps 7;

QY 32 RERAMSGKCPVMHGGNT---STGTSNKDWPEGLNLDILHQODRKSDDPDPDFNYREYR 88
Db 4 QNRQNAACP-FHGSVTNOSNRTTNKDWPNOLNLSILHQDRKNPHDEEFNYAEFQ 62
QY 89 KLDFDALKDVHVALMDSQEWMPADWGHYGGLMIRMAHWSAGTYRIADGCGGGTGSQRF 148
Db 63 KLDYWALKEDLRKLTMESQDWPADYGHYGPLFIRMAHWSAGTYRIGDGRGGASTGTQRF 122
QY 149 APLNSPDPNVSLDKARLLWPDKKKYGNKISWADLMILAGTVAYESMGLPAYGFSGRVD 208
Db 123 APLNSPDPNANDKARLLWPDKKKYGNKISWADLFILAGNVAIESMGGKTFGFGGRVD 182
QY 209 IWEPEKDIYWGDEKEWLAPSDERY-GDVNKPETMENPLAAVQMGILYVNPPEGVNGHPDPL 267
Db 183 VHPEDVYWGSEKWL---SERYSGD---RELENPLAAVQMGILYVNPPEGDGFDPK 236
QY 268 RTAQVLETFARMAMNDEKTAALTAGGHTVGNCHGNASALAPDPKASDVENQGLGWN 327
Db 237 AAARDIRETFRRMGNDEETVALIAGGHTFGAHGAGATHVGPPEAAPIEAQGLGWS 296
QY 328 PNMGGKASNAVTSIGEGAMTTPKFDMGYFDLLFGYNWELKKSPPAGAHWPEIDIKKEN 387
Db 297 SYGKGKGSITISGIEGAMTPTPTQMDTSYFDMLFGYDWMWLTSPAGAWQMAVDPEKD 356
QY 388 KPVADSDPSIRHNPIMTDADMAIKVNPYRAICEKFMADPEYFKKTFKAWFKLTHRDIG 447
Db 357 LAPDAEDPSKVPYTMWMTDLALRDFDPEYKIAARHQNPQPEEFAFAFAWFKLTHRDG 416
QY 448 PKRSYRGPVPAEDLIWQDPIAGNTDYCE---EVVKQIAOSGLSISGMVSTAWDSAT 504
Db 417 PKTRYLGVPVPEDEFTWQDPIPEVDYELTEABEIEIKAKILNSGLTVSELVKTANASAT 476
QY 505 YRGSMDRGANGARITLAPQNEWQGNPEPERLAKVLSYEQISND--TGASIDAVIVLAGS 562
```

```
Db 477 FRNSDRGGGANGARIRLAPQKQDWEVNEPERLAKVLSYVEDIQRELPKVSIADLVLGGS 536
QY 563 VGTAKAAGYDVRVPEFLKGRGDATAEMTDADSFAPLEPLADGFRNQKQEVVYKPEEM 622
Db 537 AAVEKAARDAGFDVKVPPFFGRCGDATQEQDQDVSFAVLEPFADGFRNYQKQESVPPPEL 596
QY 623 LLDRAQLMGLTGPEMTVLLGGMVLTGNTYGGTGHGVTDCGQLTNDFFVNLNLDGMSWK 682
Db 597 LVDKAQLGLGTAPMTVVLGGLRVLCANYRDLPHGVFTDRIGVLTNDFFVNLNLDGMYEW 656
QY 683 PVGSNAYEIRDRKTGAVKWTASRVDLVFGSLSLLRSYAEEYAQDDNGEKFVRDFVAATK 742
Db 657 PTDSGIYEIRDRKTGEVRWTATRVDLIFGNSILRSYAEEYAQDDNGEKFVRDFINAWK 716
QY 743 VMNADRFV 751
Db 717 VMNADRFDL 725

RESULT 10
T43878
catalase (EC 1.11.1.6) [imported] - synthetic
C:Species: synthetic
A:Note: cat gene engineered and expressed in Bacillus stearothermophilus
C:Date: 24-Mar-2000 #sequence_revision 24-Mar-2000 #text_change 23-Mar-2001
C:Accession: T43878
R:Matsumura, T.; Miyai, K.; Trakulnaleaamsai, S.; Yomo, T.; Shima, Y.; Miki, S.; Yamamoto,
Nat. Biotechnol. 17, 58-61, 1998
A:Title: Evolutionary molecular engineering by random elongation mutagenesis.
A:Reference number: Z22644; MUID:99116998; PMID:9920270
A:Accession: T43878
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-751 <MAT>
A:Cross-references: EMBL:AB020090; NID:94140087; PIDN:BAA37002.1; PID:94140088
C:Comment: For the wild-type sequence, see PIR:JS0520.
C:Genetics:
A:Gene: cat
C:Keywords: oxidoreductase

Query Match 61.3%; Score 2489; DB 4; Length 751;
Best Local Similarity 64.2%; Pred. No. 3.2e-164;
Matches 468; Conservative 85; Mismatches 160; Indels 16; Gaps 7;

QY 32 RERAMSGKCPVMHGGNT---STGTSNKDWPEGLNLDILHQODRKSDDPDPDFNYREYR 88
Db 4 QNRQNAACP-FHGSVTNOSNRTTNKDWPNOLNLSILHQDRKNPHDEEFNYAEFQ 62
QY 89 KLDFDALKDVHVALMDSQEWMPADWGHYGGLMIRMAHWSAGTYRIADGCGGGTGSQRF 148
Db 63 KLDYWALKEDLRKLTMESQDWPADYGHYGPLFIRMAHWSAGTYRIGDGRGGASTGTQRF 122
QY 149 APLNSPDPNVSLDKARLLWPDKKKYGNKISWADLMILAGTVAYESMGLPAYGFSGRVD 208
Db 123 APLNSPDPNANDKARLLWPDKKKYGNKISWADLFILAGNVAIESMGGKTFGFGGRVD 182
QY 209 IWEPEKDIYWGDEKEWLAPSDERY-GDVNKPETMENPLAAVQMGILYVNPPEGVNGHPDPL 267
Db 183 VHPEDVYWGSEKWL---SERYSGD---RELENPLAAVQMGILYVNPPEGDGFDPK 236
QY 268 RTAQVLETFARMAMNDEKTAALTAGGHTVGNCHGNASALAPDPKASDVENQGLGWN 327
Db 237 AAARDIRETFRRMGNDEETVALIAGGHTFGAHGAGATHVGPPEAAPIEAQGLGWS 296
QY 328 PNMGGKASNAVTSIGEGAMTTPKFDMGYFDLLFGYNWELKKSPPAGAHWPEIDIKKEN 387
Db 297 SYGKGKGSITISGIEGAMTPTPTQMDTSYFDMLFGYDWMWLTSPAGAWQMAVDPEKD 356
QY 388 KPVADSDPSIRHNPIMTDADMAIKVNPYRAICEKFMADPEYFKKTFKAWFKLTHRDIG 447
Db 357 LAPDAEDPSKVPYTMWMTDLALRDFDPEYKIAARHQNPQPEEFAFAFAWFKLTHRDG 416
```

```
QY 448 PKSRIGPEVPAEDLIWQDPIAGNTDYCE---EVVKOKIAQSGLSISSEMVSTAWDSART 504
Db 417 PKTRYLGPEVPEKEDIWQDPIPEVDYELTEIEIEIKAKILNSGLTVSELVKTAWASAST 476
QY 505 YRGSMDRGANGARLRLAPQNEWQNEPERLAKVLSVYEIOISAD--TCASADIVLVLAGS 562
Db 477 FNSDKRGANGARLRLAPQKDWNEPERLAKVLSVYEDIQRELPKKVSIADLVLVGGSS 536
QY 563 VGIEKAAKAAGYDVVRVPLKGRGDATAEMTDADSPAPLEPLADGFRNMQKREYVVKPEEM 622
Db 537 AAVEKAARDAGFDVKVPFPFGRGDATQEQTDVESFAVLEPPADGFRNFKQKQESVYPEEL 596
QY 623 LLDRQAQLMGLTGPENTVILGMRVLGTYNGGTHKGVTDCGQLNDFPNVLTDMGNSWK 682
Db 597 LVDKAQLLGLTAPETNVLVGLLRVLGANYRDLPHGVFTDRIGLVTFNDFVNLDMNYEW 656
QY 683 PVGSNAYEIRDRKTGAVKWTASRVDLVFGSNLLRSYAEVYAOADDNGEKFVRDFVAAWTK 742
Db 657 PTDSGIYEIRDRKTGEVTRATRVLDLFGSNILRSYAEFYAODDNOEKFVRDFINAWVK 716
QY 743 VNNADRFDV 751
Db 717 VNNADREFDL 725

RESULT 11
T43880
catalase (EC 1.11.1.6) [imported] - synthetic
C:Species: synthetic
A:Note: cat gene engineered and expressed in Bacillus stearothermophilus
C:Date: 24-Mar-2000 #sequence_revision 24-Mar-2000 #text_change 23-Mar-2001
C:Accession: T43880
R:Matsuura, T.; Miyai, K.; Trakulnaleaamsai, S.; Yomo, T.; Shima, Y.; Miki, S.; Yamamoto
Nat. Biotechnol. 17, 58-61, 1998
A:Title: Evolutionary molecular engineering by random elongation mutagenesis.
A:Reference number: Z22644; MUID:99116998; PMID:9920270
A:Accession: T43880
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-751 <MAT>
A:Cross-references: EMBL:AB020091; NID:g4140089; PIDN:BAA37003.1; PID:g4140090
C:Comment: For the wild-type sequence, see PIR:JS0520.
C:Genetics:
A:Gene: cat
C:Keywords: oxidoreductase

Query Match 61.3%; Score 2489; DB 4; Length 751;
Best Local Similarity 64.2%; Pred. No. 3.2e-164;
Matches 468; Conservative 85; Mismatches 160; Indels 16; Gaps 7;

QY 32 RERAMSGKCPVMHGGT---STGTGNKDWPEGLNLDILHQODRKSDDPMDPNFYREYR 88
Db 4 QNRQNAACQP-FHGSVTNQSSNRTTKDWPNQLNLSILHQHDKTNPHDEEFNVAEEFQ 62
QY 89 KLDFFALKKDVALMTDSQEWPPADGWGHYGGLMIRMAHWSAGTYRIADGRGGGTGSQRF 148
Db 63 KLDYWALKEDLRKLMTESQDWWPADYGHYGLPIFIRMAHWSAGTYRIGDGRGGASTGTQRF 122
QY 149 APLNSWPDNVSLDKARRLLWPDKKYGKNSIADLMLAGTVAYESMGLPAYGFSFGYRVD 208
Db 123 APLNSWPDNANLDKARRLLWPDKKYGKNSIADLMLAGTVAYESMGGKTI GFGGGRVD 182
QY 209 IWEKDIYWGDEKEWLAPSDERY-GDVNKPETMENPLAAVQMGIIYVNPVGNHPDPL 267
Db 183 VHPREDYVWGSEKWL- --SERYS- ---RELENPLAAVQMGIIYVNPVGNHPDPL 236
QY 268 RTAQOVLFTFARMANDEKTAALTAGGHTVGNCHGNASALAPDKASDVENQGLGWGN 327
Db 237 AAARDITFRFRMGNDEETVALLAGGHTFGKAGAGPATHVGPPEAAPIEAQGLGWTS 296
QY 328 PNMQKASNAVTSIGECANTPTKFDNCFYDILLFCYNWELKSPAGAHWEPIIDIKKEN 387
Db 297 SYGKGKGSDTITSGIEGAWTPTTQWDTSYFDMLFGYDWMWLT KSPAGAWQMAVDPEKD 356
```

```
QY 388 KPVDADEPSIRHNPIMTDADMAIKVNPYTRAIKCEKFMADPEYFKKTFARAWFKLTHRDIG 447
Db 357 LAPDAEDPSKKVPTWMTTDLALRFDPEVEKIFARFHQNPPEFAFAFAWFKLTHRDIG 416
QY 448 PKSRIGPEVPAEDLIWQDPIAGNTDYCE---EVVKOKIAQSGLSISSEMVSTAWDSART 504
Db 417 PKTRYLGPEVPEKEDIWQDPIPEVDYELTEIEIEIKAKILNSGLTVSELVKTAWASAST 476
QY 505 YRGSMDRGANGARLRLAPQNEWQNEPERLAKVLSVYEIOISAD--TCASADIVLVLAGS 562
Db 477 FNSDKRGANGARLRLAPQKDWNEPERLAKVLSVYEDIQRELPKKVSIADLVLVGGSS 536
QY 563 VGIEKAAKAAGYDVVRVPLKGRGDATAEMTDADSPAPLEPLADGFRNMQKREYVVKPEEM 622
Db 537 AAVEKAARDAGFDVKVPFPFGRGDATQEQTDVESFAVLEPPADGFRNFKQKQESVYPEEL 596
QY 623 LLDRQAQLMGLTGPENTVILGMRVLGTYNGGTHKGVTDCGQLNDFPNVLTDMGNSWK 682
Db 597 LVDKAQLLGLTAPETNVLVGLLRVLGANYRDLPHGVFTDRIGLVTFNDFVNLDMNYEW 656
QY 683 PVGSNAYEIRDRKTGAVKWTASRVDLVFGSNLLRSYAEVYAOADDNGEKFVRDFVAAWTK 742
Db 657 PTDSGIYEIRDRKTGEVTRATRVLDLFGSNILRSYAEFYAODDNOEKFVRDFINAWVK 716
QY 743 VNNADRFDV 751
Db 717 VNNADREFDL 725

RESULT 12
T43881
catalase (EC 1.11.1.6) [imported] - synthetic
C:Species: synthetic
A:Note: cat gene engineered and expressed in Bacillus stearothermophilus
C:Date: 24-Mar-2000 #sequence_revision 24-Mar-2000 #text_change 23-Mar-2001
C:Accession: T43881
R:Matsuura, T.; Miyai, K.; Trakulnaleaamsai, S.; Yomo, T.; Shima, Y.; Miki, S.; Yamam
Nat. Biotechnol. 17, 58-61, 1998
A:Title: Evolutionary molecular engineering by random elongation mutagenesis.
A:Reference number: Z22644; MUID:99116998; PMID:9920270
A:Accession: T43881
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-751 <MAT>
A:Cross-references: EMBL:AB020099; NID:g4140105; PIDN:BAA37011.1; PID:g4140106
C:Comment: For the wild-type sequence, see PIR:JS0520.
C:Genetics:
A:Gene: cat
C:Keywords: oxidoreductase

Query Match 61.3%; Score 2489; DB 4; Length 751;
Best Local Similarity 64.2%; Pred. No. 3.2e-164;
Matches 468; Conservative 85; Mismatches 160; Indels 16; Gaps 7;

QY 32 RERAMSGKCPVMHGGT---STGTGNKDWPEGLNLDILHQODRKSDDPMDPNFYREYR 88
Db 4 QNRQNAACQP-FHGSVTNQSSNRTTKDWPNQLNLSILHQHDKTNPHDEEFNVAEEFQ 62
QY 89 KLDFFALKKDVALMTDSQEWPPADGWGHYGGLMIRMAHWSAGTYRIADGRGGGTGSQRF 148
Db 63 KLDYWALKEDLRKLMTESQDWWPADYGHYGLPIFIRMAHWSAGTYRIGDGRGGASTGTQRF 122
QY 149 APLNSWPDNVSLDKARRLLWPDKKYGKNSIADLMLAGTVAYESMGLPAYGFSFGYRVD 208
Db 123 APLNSWPDNANLDKARRLLWPDKKYGKNSIADLMLAGTVAYESMGGKTI GFGGGRVD 182
QY 209 IWEKDIYWGDEKEWLAPSDERY-GDVNKPETMENPLAAVQMGIIYVNPVGNHPDPL 267
Db 183 VHPREDYVWGSEKWL- --SERYS- ---RELENPLAAVQMGIIYVNPVGNHPDPL 236
QY 268 RTAQOVLFTFARMANDEKTAALTAGGHTVGNCHGNASALAPDKASDVENQGLGWGN 327
Db 297 SYGKGKGSDTITSGIEGAWTPTTQWDTSYFDMLFGYDWMWLT KSPAGAWQMAVDPEKD 356
```

Db 237 AARDIRETRFRMGNDDEETVALIAGGHTFGKAHAGGATHVGPPEAPAEIAQGLGWS 296
 QY 328 PMQKASNAVTSIGEGAWTNTPTKFDGMYFDLLFGYNWELKSPAGAHHPIDIKKEN 387
 Db 297 SYGKKGSDTITSGIEGAWTPTPTQWDTSYFDMLECYDMLTKSPAGAWQWMAVDPDEKD 356
 QY 388 KPVDSASDSIRHNPTMTDADMAIKVNPYRAICEKFMADPEYFKTKFAKAFKLTDRDLG 447
 Db 357 LAPDAEDPSKVPVTMMTDLALRDFPEYKIAARFHFONPEFAFAFAFAFWKLTDRDWG 416
 QY 448 PKSYIGPEVPAEDLWQDPIPAAGNTDYCE---EYVKOKIAQSGLSISEMYSTANDSART 504
 Db 417 PKTRYLGPEVKEDFIWQDPIPEVDYELTEAIEIEIKAKILNSGLTVSELVKTAWASAST 476
 QY 505 YRGSMDRGANGARIRLAPQWNOGNEPERLAKVLSVYEIQISAD--TCGASIAADVIVLAGS 562
 Db 477 FRNSDKRGANGARIRLAPQKDWMEVNEPERLAKVLSVYEDIQRELPKKVSIADLIVLGS 536
 QY 563 VGIEKAAKAGDVVRPFLKRGDATAETDADSPAPLEPLADGFRNKKQKYEYVYKPEEM 622
 Db 597 LVDRQAQLGLTAPENTVLVGLRVLGANYRDLPHGVFTDRIGVLTNDFVNLDMNYEW 556
 QY 683 PVGSNAYEIRDRKTKGAVKWTASRVDLVFGSNLSRSYAENVYAQDDNGKFKVRDFVAAWTK 742
 Db 657 PTDSGIYEIRDRKTKGEVWTATRVDLIFGNSILRSYAENVYAQDDNQEKFKVRDFINAWVK 716
 QY 743 YMNADRDV 751
 Db 717 YMNADRDV 725

RESULT 13

T43877
 catalase (EC 1.11.1.6) [imported] - synthetic
 C:Species: synthetic
 A:Note: cat gene engineered and expressed in *Bacillus stearothermophilus*
 C:Date: 24-Mar-2000 #sequence_revision 24-Mar-2000 #text_change 23-Mar-2001
 C:Accession: T43877
 R:Matsumura, T.; Miyai, K.; Trakulnaleamsai, S.; Yomo, T.; Shima, Y.; Miki, S.; Yamamoto
 Nat. Biotechnol. 17, 58-61, 1998
 A:Title: Evolutionary molecular engineering by random elongation mutagenesis.
 A:Reference number: 22644; MUID:99116998; PMID:9920270
 A:Accession: T43877
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-752 <MAT>
 A:Cross-references: EMBL:AB020089; NID:g4140085; PIDN:BA37001.1; PID:g4140086
 C:Comment: For the wild-type sequence, see PIR:JS0520.
 C:Genetics:
 A:Gene: cat
 C:Keywords: oxidoreductase

Query Match 61.3%; Score 2489; DB 4; Length 752;
 Best Local Similarity 64.2%; Pred. No. 3.2e-164;
 Matches 468; Conservative 85; Mismatches 160; Indels 16; Gaps 7;

QY 32 RERAMSKCPVMHGNT---STGTSNKDWPEGLNLDILHQDRKSDPDPDFNYREYR 88
 Db 4 QNRQNAQCP-FHGSVTNQSSNTTNKDWPNQNLNLSILHQRKTNPHDEEFYAEFFQ 62
 QY 89 KLDFALKKDVAHMLTDSQEWNPADWGHYGLGIMTIRMAHWSAGTYRIADRGGGTGSORF 148
 Db 63 KLDYWALKEDLRKLMTESQDWWNPADYGHYGLFLIRMAHWSAGTYRIADRGGGTGSORF 122
 QY 149 APLNSWPDNLSOKARLLPIPKKKYGNKISWADLMLAGTVAYESMGLPAYGSGFRVD 208
 Db 123 APLNSWPDNLSOKARLLPIPKKKYGNKISWADLMLAGTVAYESMGLPAYGSGFRVD 182
 QY 209 IWEKDIYNGDEKEWLAQSDERY-GDVNKPETMENPAAVQMGGLIYVNGYNGHPDPL 267

Db 183 VHPBEDYVWGSEKELA---SERYSGD-----RELENPAAVQMGGLIYVNGYNGYDPDKPDK 236
 QY 268 RFAQOVLETFAWAMNDEKTAALTAGGHTVGNCHGNASALAPDKPASKDVNOGLGWN 327
 Db 237 AARDIRETRFRMGNDDEETVALIAGGHTFGKAHAGGATHVGPPEAPAEIAQGLGWS 296
 QY 328 PMQKASNAVTSIGEGAWTNTPTKFDGMYFDLLFGYNWELKSPAGAHHPIDIKKEN 387
 Db 297 SYGKKGSDTITSGIEGAWTPTPTQWDTSYFDMLECYDMLTKSPAGAWQWMAVDPDEKD 356
 QY 388 KPVDSASDSIRHNPTMTDADMAIKVNPYRAICEKFMADPEYFKTKFAKAFKLTDRDLG 447
 Db 357 LAPDAEDPSKVPVTMMTDLALRDFPEYKIAARFHFONPEFAFAFAFAFWKLTDRDWG 416
 QY 448 PKSYIGPEVPAEDLWQDPIPAAGNTDYCE---EYVKOKIAQSGLSISEMYSTANDSART 504
 Db 417 PKTRYLGPEVKEDFIWQDPIPEVDYELTEAIEIEIKAKILNSGLTVSELVKTAWASAST 476
 QY 505 YRGSMDRGANGARIRLAPQWNOGNEPERLAKVLSVYEIQISAD--TCGASIAADVIVLAGS 562
 Db 477 FRNSDKRGANGARIRLAPQKDWMEVNEPERLAKVLSVYEDIQRELPKKVSIADLIVLGS 536
 QY 563 VGIEKAAKAGDVVRPFLKRGDATAETDADSPAPLEPLADGFRNKKQKYEYVYKPEEM 622
 Db 597 LVDRQAQLGLTAPENTVLVGLRVLGANYRDLPHGVFTDRIGVLTNDFVNLDMNYEW 556
 QY 683 PVGSNAYEIRDRKTKGAVKWTASRVDLVFGSNLSRSYAENVYAQDDNGKFKVRDFVAAWTK 742
 Db 657 PTDSGIYEIRDRKTKGEVWTATRVDLIFGNSILRSYAENVYAQDDNQEKFKVRDFINAWVK 716
 QY 743 YMNADRDV 751
 Db 717 YMNADRDV 725

RESULT 14

T43882
 catalase (EC 1.11.1.6) [imported] - synthetic
 C:Species: synthetic
 A:Note: cat gene engineered and expressed in *Bacillus stearothermophilus*
 C:Date: 24-Mar-2000 #sequence_revision 24-Mar-2000 #text_change 23-Mar-2001
 C:Accession: T43882
 R:Matsumura, T.; Miyai, K.; Trakulnaleamsai, S.; Yomo, T.; Shima, Y.; Miki, S.; Yamamoto
 Nat. Biotechnol. 17, 58-61, 1998
 A:Title: Evolutionary molecular engineering by random elongation mutagenesis.
 A:Reference number: 22644; MUID:99116998; PMID:9920270
 A:Accession: T43882
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-753 <MAT>
 A:Cross-references: EMBL:AB020119; NID:g4140145; PIDN:BA37031.1; PID:g4140146
 C:Comment: For the wild-type sequence, see PIR:JS0520.
 C:Genetics:
 A:Gene: cat
 C:Keywords: oxidoreductase

Query Match 61.3%; Score 2489; DB 4; Length 753;
 Best Local Similarity 64.2%; Pred. No. 3.2e-164;
 Matches 468; Conservative 85; Mismatches 160; Indels 16; Gaps 7;

QY 32 RERAMSKCPVMHGNT---STGTSNKDWPEGLNLDILHQDRKSDPDPDFNYREYR 88
 Db 4 QNRQNAQCP-FHGSVTNQSSNTTNKDWPNQNLNLSILHQRKTNPHDEEFYAEFFQ 62
 QY 89 KLDFALKKDVAHMLTDSQEWNPADWGHYGLGIMTIRMAHWSAGTYRIADRGGGTGSORF 148
 Db 63 KLDYWALKEDLRKLMTESQDWWNPADYGHYGLFLIRMAHWSAGTYRIADRGGGTGSORF 122
 QY 149 APLNSWPDNLSOKARLLPIPKKKYGNKISWADLMLAGTVAYESMGLPAYGSGFRVD 208
 Db 123 APLNSWPDNLSOKARLLPIPKKKYGNKISWADLMLAGTVAYESMGLPAYGSGFRVD 182
 QY 209 IWEKDIYNGDEKEWLAQSDERY-GDVNKPETMENPAAVQMGGLIYVNGYNGHPDPL 267

149 APLNSPDNVSIDKARLLWPKKKYGNKISWADLMILAGTVAYESMGLPAYGFSGRVD 208
 123 APLNSPDNANLDKARLLWPKKKYGNKISWADLMILAGTVAYESMGLPAYGFSGRVD 182
 209 IWEPEKDIYWGDEKEWLPASDERY -GDVANKPETMENPLAAVOMGLIYVNPGEVNGHPDPL 267
 183 VHPEDVYWGSEKELA -SERYSQD -RELENPLAAVOMGLIYVNPGEVNGHPDPR 236
 268 RTAQVLETFARMAMNDEKTAALTAGHTVGNCHGNSALAPDPKASDVENQGLWGN 327
 237 AAARDIRETFRRMNMDEETVALIAGHTFGKAGGAGPATHVGPPEAAPIEAQGLWIS 296
 328 PNMOKASNAVYSGIGANTTNPTEKDMGYEDLLFGYNELKSPAGAHNWEPIIDIKEN 387
 297 SYGKRGSGTITSGIGANTTPTQWDTSYFDMFLGYDMWLTSPAGAWQMAVDPEK 356
 388 KPVDAADSPSIRHNPIMTDADMAIKVNPVTRAIKEKPMADPEVFKTKAKWPKLTHRDLG 447
 357 LAPDAEDPSKVPYTMNTDALARFDPPEYKTRRRHQNEPEEFAFAFAWFKLTHRDG 416
 448 PKSRITGPEVPAEDLIWQDPIIPAGNTDYCE -EYVKQKIAQSGLSISEMSTWASDART 504
 417 PKTRYLGPEVPEKDFWQDPIPEVDYVELTEAEIEIKAKILNSGLTVSELVKTWASAST 476
 505 YRGSDMRGGANGARIRLAFQNEQNEPERLAKVLSYEQISAD -TGASIAVDVIVLAGS 562
 477 FRNSDKRGGANGARIRLAFQNEQNEPERLAKVLSYEDIQRELPKVSIADLIYLGGS 536
 563 VGTAKAAAGYDVRVPFTLKGRCDAETADSPAPLEPLADGPRNWKKEVVKPEEM 622
 537 AAVEKARADGDFVKKVFPFGRGDQATQEQTDVESFAVLEPFAFGFRYKQESVSPPEEL 596
 623 LLDRAQLMGLTGPEMVLVGLGRVLTGNTGGTKHGVTDCBGLTNDFFVNLTDGNSWK 682
 597 LVDKAQLGLTAPEMVLVGLRVLGANYRDLPHGVFTDRIGVLTNDFFVNLDMVYEW 656
 683 PVGSNAYEIRDRGTGAVKKTASVDVLFVGSNLSLRSAEYVAQDDNGEKFVDFVAWTK 742
 657 PTDGSGIYEIRDRGTGAVKKTASVDVLFVGSNLSLRSAEYVAQDDNGEKFVDFVAWTK 716
 743 VHNADRFV 751
 717 VHNADRFDL 725

RESULT 15

F82584

catalase/peroxidase XF2232 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 24-Aug-2001

C:Accession: F82584

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: AB2515; MUID:20365717; PMID:10910347

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: F82584

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-781 <SIM>

A:Cross-references: GB:AE004035; GB:AE003849; NID:99107371; PIDN:AAF85031.1; GSPDB:GN001

R:Experimental source: strain 9a5c

A:Authors: A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A

as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, H

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir

M.: Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
 A:Reference number: A59328

A:Contents: annotation

C:Genetics:

C:Gene: XF2232

C:Superfamily: catalase HPI

C:Keywords: heme; iron; metalloprotein

F:126/Active site: His (distal axial ligand) #status predicted

F:313/Binding site: heme iron (His) (proximal axial ligand) #status predicted

F:364/Active site: Trp (tryptophyl radical intermediate) #status predicted

Query Match 56.2%; Score 2279; DB 2; Length 781;

Best Local Similarity 56.8%; Pred. No. 1.2e-149;

Matches 431; Conservative 100; Mismatches 176; Indels 54; Gaps 8;

QY 40 CPVHGGTST-----GTSNKNMPEGLNLDILHQDQKSDPMDPDNYRE 85
 DB 25 CPVNRASSPETTSVAVKCPNKTAVENTHKNQWPNQLRVDLLHQHNSKNPLGETFDYAK 84
 QY 86 EVKRLDFDALKDVHALMTDSQEMWADMGHYGGLMIRMAWHSAGTYRIADGGGGTGS 145
 DB 85 EFQKLDYAALKRDLHALMTDSQEMWADMGHYGGLFIRMAWHSAGTYRIADGGGGTGS 144
 QY 146 ORFAPLNSPDNVSIDKARLLWPKKKYGNKISWADLMILAGTVAYESMGLPAYGFSFG 205
 DB 145 ORFAPLNSPDNVSIDKARLLWPKKKYGNKISWADLMILAGTVAYESMGLPAYGFSFG 204
 QY 206 RVDIWEPEKDIYWGDEKEWLPASDERYGVNKK-----PETWENPL 245
 DB 205 RVDIWEPEKDIYWGDEKEWLPASDERYGVNKK-----PETWENPL 245
 QY 246 AAYQMGGLIYVNPGEVNGHPDPLTAQVLETFAARMANDEKTAALTAGHTVGNCHGN 305
 DB 264 AAYQMGGLIYVNPGEVNGHPDPLTAQVLETFAARMANDEKTAALTAGHTVGNCHGN 305
 QY 306 ASALADPDKASDVENQNEGNGNPMOGKASNAVTSIGEGAWTNPTRKFDGMYDGLFGYN 365
 DB 324 ADVNGPEPEAGELEQOGLGWHNRFSGKAGDTITSGLEVTWTKTPTQWSDNDFEHLFGYE 383
 QY 366 WELKKGSPAGAHNWEPIIDIKENKPV--DASDPSIRHNPIMTDADMAIKVNPVTRAIKEKF 423
 DB 384 WELKKGSPAGAHNWEPIIDIKENKPV--DASDPSIRHNPIMTDADMAIKVNPVTRAIKEKF 423
 QY 424 MADPEYFKKTFKAKWPKLTHRDGLPKSRYSYIGPEVPAEDLIWQDPIIPAGNTDYCEE---V 479
 DB 440 HAHPDQFADVFARAWFKLHMDMGPRVRYLGPVPEELIWDQVVPKSHVLDQAQLLA 499
 QY 480 VKOKIASGSLISEMSTWASDARTYRGSDMRGGANGARIRLAPQNEWQNEPERLAKVL 539
 DB 500 LKOKIASGSLISEMSTWASDARTYRGSDMRGGANGARIRLAPQNEWQNEPERLAKVL 539
 QY 540 SVYEQISADTGA-----SIADVIVLAGSVGIEKAAAGYDVRVPPELKGRCGDATWMT 592
 DB 560 ETUERRVQTEFNAAGDKRISLADLIYVLAGGVGVEQAARAGIVVEVFPVFPERTDAQEQT 619
 QY 593 DADSFAPLEPLADGFRNWRKQYVYKPEEMLLDRAQLMGLTGPMTVLLGGMYVLTGNYG 652
 DB 620 DVSSFAPLEPLADGFRNWRKQYVYKPEEMLLDRAQLMGLTGPMTVLLGGMYVLTGNYG 652
 QY 653 GTHGVFTDCGGLTNDFFVNLTDGMSNWKPVGS--NAYEIRKKGKAVKWTASRVLDLVF 710
 DB 680 GVKHGVTDRGLTSLNSDFFINLDMGTENAPVSKERHVFEGDRRTGTGLKWTGTVDLVF 739
 QY 711 GSNLSLSRYAEVTAADNDGKEKFRVDFVAATKVMNADRPDV 751
 DB 740 GSNALLRALAEFYAAVDAQKEKFRVDFVAATKVMNADRPDV 751

Search completed: October 7, 2003, 19:46:43

Job time : 32.6575 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 7, 2003, 19:38:38 ; Search time 75.4005 Seconds
(without alignments)
2577.085 Million cell updates/sec

Title: US-09-884-889-6
Perfect score: 4058
Sequence: 1 MNASADLHSLQRCRAF.....RDFVAANTKVMNDRFDVAS 753

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2920	72.0	721	Q9ZGM4	Q9ZGM4 legionella
2	2919	71.9	724	Q9KRS6	Q9KRS6 vibrio chol
3	2870	70.7	720	Q9R6S9	Q9R6S9 synecococc
4	2854	70.3	723	Q8D951	Q8D951 vibrio vuln
5	2820	69.5	731	Q8U714	Q8U714 agrobacteri
6	2817	69.4	723	Q9R708	Q9R708 agrobacteri
7	2802.5	69.1	728	Q8E981	Q8E981 shewanella
8	2795	68.9	754	P73911	P73911 synecocyst
9	2731	67.3	720	Q55110	Q55110 synecococc
10	2514	62.0	735	Q9KEE5	Q9KEE5 bacillus ha
11	2489	61.3	743	Q9S5K5	Q9S5K5 bacillus st
12	2489	61.3	743	Q9S5R0	Q9S5R0 bacillus st
13	2489	61.3	744	Q9S5N2	Q9S5N2 bacillus st
14	2489	61.3	744	Q9S5K7	Q9S5K7 bacillus st
15	2489	61.3	745	Q9S5Q5	Q9S5Q5 bacillus st
16	2489	61.3	746	Q9S5N4	Q9S5N4 bacillus st

17	2489	61.3	748	2	Q9S5L8	Q9S5L8 bacillus st
18	2489	61.3	749	2	Q9S5P6	Q9S5P6 bacillus st
19	2489	61.3	750	2	Q9S5Q6	Q9S5Q6 bacillus st
20	2489	61.3	750	2	Q9S5Q7	Q9S5Q7 bacillus st
21	2489	61.3	751	2	Q9S5P7	Q9S5P7 bacillus st
22	2489	61.3	751	2	Q9S5L5	Q9S5L5 bacillus st
23	2489	61.3	751	2	Q9S5N9	Q9S5N9 bacillus st
24	2489	61.3	751	2	Q9S5P1	Q9S5P1 bacillus st
25	2489	61.3	751	2	Q9S5L1	Q9S5L1 bacillus st
26	2489	61.3	751	2	Q9S5M4	Q9S5M4 bacillus st
27	2489	61.3	751	2	Q9S5M1	Q9S5M1 bacillus st
28	2489	61.3	751	2	Q9S5Q8	Q9S5Q8 bacillus st
29	2489	61.3	751	2	Q9S5P9	Q9S5P9 bacillus st
30	2489	61.3	751	2	Q9S5N0	Q9S5N0 bacillus st
31	2489	61.3	751	2	Q9S5N3	Q9S5N3 bacillus st
32	2489	61.3	751	2	Q9S5M0	Q9S5M0 bacillus st
33	2489	61.3	751	2	Q9S5K4	Q9S5K4 bacillus st
34	2489	61.3	751	2	Q9S5P2	Q9S5P2 bacillus st
35	2489	61.3	751	2	Q9S5Q0	Q9S5Q0 bacillus st
36	2489	61.3	751	2	Q9S5N6	Q9S5N6 bacillus st
37	2489	61.3	751	2	Q9S5L7	Q9S5L7 bacillus st
38	2489	61.3	751	2	Q9S5Q3	Q9S5Q3 bacillus st
39	2489	61.3	751	2	Q9S5M7	Q9S5M7 bacillus st
40	2489	61.3	751	2	Q9S5M8	Q9S5M8 bacillus st
41	2489	61.3	751	2	Q9S5L6	Q9S5L6 bacillus st
42	2489	61.3	751	2	Q9S5L9	Q9S5L9 bacillus st
43	2489	61.3	751	2	Q9S5L3	Q9S5L3 bacillus st
44	2489	61.3	751	2	Q9S5Q1	Q9S5Q1 bacillus st
45	2489	61.3	751	2	Q9S5Q4	Q9S5Q4 bacillus st

ALIGNMENTS

RESULT 1

Q9ZGM4 PRELIMINARY; PRT; 721 AA.
AC Q9ZGM4;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Catalase-peroxidase KatB (EC 1.11.1.6) (Peroxidase/catalase).
GN KATB.
OS Legionella pneumophila.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Legionellaceae; Legionella.
OX NCBI_TaxID=446;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JR32;
RX MEDLINE=98440428; PubMed=9765568;
RA Bandyopadhyay P., Steinman H.M.;
RT "Legionella pneumophila catalase-peroxidases: cloning of the katB gene and studies of KatB function."
RL J. Bacteriol. 180:5369-5374(1998).
CC -|- CATALYTIC ACTIVITY: 2 H(2)O(2) -> O(2) + 2 H(2)O.
CC -|- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL PEROXIDASE/CATALASE SUBFAMILY.
CC EMBL: AF078110; AAC64361.1; .
DR HSSP: P48534; IAPX.
DR InterPro: IPR000763; Bac_ctase/prase.
DR InterPro: IPR002016; Peroxidase.
DR Pfam: PF00141; peroxidase; 1.
DR PRINTS: PR00458; PEROXIDASE.
DR TIGRFS: TIGR00198; cat_per_HPI; 1.
DR PROSITE: PS00436; PEROXIDASE_2; 1.
KW Heme; Hydrogen peroxide; iron; Oxidoreductase; Peroxidase.
SQ SEQUENCE 721 AA; 80499 MW; D96347BDAD0B719C CRC64;

Query Match 72.0%; Score 2920; DB 2; Length 721;
Best Local Similarity 73.3%; Pred. No. 1.8e-192;
Matches 531; Conservative 77; Mismatches 106; Indels 10; Gaps 3;

```
QY 36 MSGK-----CPVHGGTSTGTSTNKNKDWPEGLNLDILHQDRKSDPMDPDFNRYREVR 88
Db 1 MDGKVGSTTCPCPIHGGTSTGTSTNKNKDWPEGLNLDILHQDRKSDPMDPDFNRYREVR 60
QY 89 KLDFDALKDKVHALMTDSQEWPPADWGHYGGIMIRMAWHSAGTVRIADGRGGGTGSGORE 148
Db 61 KLDFEALKDKLHALMTDSQAWPPADWGHYGGIMIRMAWHSAGTVRIADGRGGGTGSGORE 120
QY 149 APLNSWPDNVLSDKARLLWPKKKYGNKISWADLMILAGTVAYESMGLPAYESFGFRVD 208
Db 121 APLNSWPDNVLSDKARLLWPKKKYGNKISWADLMILAGTVAYESMGLPAYESFGFRVD 180
QY 209 IWEPEKDIYWGDEKWLAPSDRYGVNKPETMENPLAAVQMGILYVNPVGVNHPDPLR 268
Db 181 IWEPEKDIYWGDEKWLAPSDRYGVNKPETMENPLAAVQMGILYVNPVGVNHPDPLR 237
QY 269 TAQOVLETFAKAMNDEKTAALTAGGHTVGNCHGNASALAPDPKASDVNOGLGWGPN 328
Db 238 TAQOVLETFAKAMNDEKTAALTAGGHTVGNCHGNASALAPDPKASDVNOGLGWGPN 297
QY 329 NMOKKASNAVTSIGEGAWTNTPTKFDGMYFDLLFGYNWELKKSPAGAHNHPIDIKKENK 388
Db 298 TTRGIRGNTSSGIEGAWTNTPTKFDGMYFDLLFGYNWELKKSPAGAHNHPIDIKKENK 357
QY 389 PVDSASDSIRHNPITMDADMAIKVNPYRAICEKFMADPEYFKTFKAKWFKLTHRDLP 448
Db 358 PVDSASDSIRHNPITMDADMAIKVNPYRAICEKFMADPEYFKTFKAKWFKLTHRDLP 417
QY 449 KSRIGYGEVPAEDLIWDPIDPAGTNDYCEEVVKQIAOSGLSISEMYSTANDSARTYRS 508
Db 418 KTRIGYGEVPAEDLIWDPIDPAGTNDYCEEVVKQIAOSGLSISEMYSTANDSARTYRS 477
QY 509 DMRGANGARIRLAPQWEGNEPERLAKVLSVYEQISADTGASIAVIVLAGSVGIEKA 568
Db 478 DMRGANGARIRLAPQWEGNEPERLAKVLSVYEQISADTGASIAVIVLAGSVGIEKA 537
QY 569 AKAAGYDVRVPEFLKGRGDATAEMTDADSEAPLEPLADGFRNWKQYVYVPEEMLLDRAQ 628
Db 538 AKAAGYDVRVPEFLKGRGDATAEMTDADSEAPLEPLADGFRNWKQYVYVPEEMLLDRAQ 597
QY 629 LMGLTGPEMTVLGGMRVLGNTYGGTKHGVFTDCEGOLTDNDFVNLTDGMSWKPVGNSA 688
Db 598 LMGLTGPEMTVLGGMRVLGNTYGGTKHGVFTDCEGOLTDNDFVNLTDGMSWKPVGNSA 657
QY 689 YEIRDRITGAVKWTATSRVDLVFGNSLLRSYAEVYAOQDNKEKFRVDFVAATKVMNADR 748
Db 658 YEIRDRITGAVKWTATSRVDLVFGNSLLRSYAEVYAOQDNKEKFRVDFVAATKVMNADR 717
QY 749 FDVA 752
Db 718 FDVA 721
```

RESULT 2

```
Q9KRS6 PRELIMINARY; PRT: 724 AA.
AC Q9KRS6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Catalase/peroxidase.
GN VCL1560.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Elsen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
```

RESULT 3

```
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
DR EMBL; AE004233; AAF94714.1; -.
DR HSSP; P00431; 1BEK.
DR TIGR; VCL1560; -.
DR InterPro; IPR000763; Bac_ctase/prase.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00141; peroxidase; 1.
DR PRINTS; PR00458; PEROXIDASE.
DR TIGRFAMs; TIGR00198; cat_per_HPI; 1.
DR PROSITE; PS00436; PEROXIDASE_2; 1.
DR Peroxidase; Complete proteome.
KW SEQUENCE 724 AA; 80650 MW; C53E123D12F6DB51 CRC64;
SQ
Query Match 71.9%; Score 2919; DB 16; Length 724;
Best Local Similarity 74.1%; Pred. No. 2.1e-192;
Matches 530; Conservative 75; Mismatches 110; Indels 0; Gaps 0;
QY 37 SGKCPVHGGTSTGTSTNKNKDWPEGLNLDILHQDRKSDPMDPDFNRYREVRKLFDFALK 96
Db 9 SGKCPVHGGTSTGTSTNKNKDWPEGLNLDILHQDRKSDPMDPDFNRYREVRKLFDFALK 68
QY 97 KDVALMTDSQEWPPADWGHYGGIMIRMAWHSAGTVRIADGRGGGTGSGORFAPLNSWPD 156
Db 69 RDLKALMTNSQEWPPADWGHYGGIMIRMAWHSAGTVRIADGRGGGTGSGORFAPLNSWPD 128
QY 157 NVSLDKARRLLWPKKKYGNKISWADLMILAGTVAYESMGLPAYESFGFRVDIWEPEKDI 216
Db 129 NANLDKARRLLWPKKKYGNKISWADLMILAGTVAYESMGLPAYESFGFRVDIWEPEKDI 188
QY 217 YWGDEKWLAPSDRYGVNKPETMENPLAAVQMGILYVNPVGVNHPDPLRQAQVLET 276
Db 189 YWGSEKWLAKSGGNSRYSGORLENPAAVMHGLIYVNPVGVNHPDPLRQAQVLET 248
QY 277 FARMAMNDEKTAALTAGGHTVGNCHGNASALAPDPKASDVNOGLGWGPNMOGKASN 336
Db 249 FARMAMNDEKTAALTAGGHTVGNCHGNASALAPDPKASDVNOGLGWGPNMOGKASN 308
QY 337 AVTSGIEGAWTNTPTKFDGMYFDLLFGYNWELKKSPAGAHNHPIDIKKENKPYDASDP 396
Db 309 TVTSGIEGAWTTHPTRWDNVYFLLSYEWOLTKSPAGAWQWEPVNIKEEDKPYDVEDPS 368
QY 397 IRHNPITMDADMAIKVNPYRAICEKFMADPEYFKTFKAKWFKLTHRDLPKSRVYQPE 456
Db 369 IRYNPMNTMDADMAIKVNPYRAICEKFMADPEYFKTFKAKWFKLTHRDLPKSRVYQPE 428
QY 457 VPAEDLIWQDPIPAAGTNDYCEEVVKQIAOSGLSISEMYSTANDSARTYRSGDRGGANG 516
Db 429 VPAEDLIWQDPIPAAGTNDYCEEVVKQIAOSGLSISEMYSTANDSARTYRSGDRGGANG 488
QY 517 ARIRLAPQWEGNEPERLAKVLSVYEQISADTGASIAVIVLAGSVGIEKAAGAAYDV 576
Db 489 ARIRLAPQWEGNEPERLAKVLSVYEQISADTGASIAVIVLAGSVGIEKAAGAAYDV 548
QY 577 RVPLKGRGDATAEMTDADSEAPLEPLADGFRNWKQYVYVPEEMLLDRAQLMGLTQPE 636
Db 549 TVPFAPGRGDATIEQTDVESFEVLEPLADGFRNWKQYVYVPEEMLLDRAQLMGLTQPE 608
QY 637 MTVLLGGMRVLGNTYGGTKHGVFTDCEGOLTDNDFVNLTDGMSWKPVGNSNAYEIRDRKT 696
Db 609 MTVLLGGMRVLGNTYGGTKHGVFTDCEGOLTDNDFVNLTDGMSWKPVGNSNAYEIRDRKT 668
QY 697 GAVKWTATSRVDLVFGNSLLRSYAEVYAOQDNKEKFRVDFVAATKVMNADRFDV 751
Db 669 GAVKWTATSRVDLVFGNSLLRSYAEVYAOQDNKEKFRVDFVAATKVMNADRFDV 723
```

Q9R6S9 Q9R6S9 PRELIMINARY; PRT; 720 AA.

AC Q9R6S9; 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Catalase-peroxidase (EC 1.11.1.6) (Peroxidase/catalase).

OS *Synechococcus* sp. (strain PCC 6301) (Anacystis nidulans).

OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.

OX NCBI_TaxID=1139;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=PCC 6301;

RX MEDLINE=20322938; PubMed=10863004;

RA Engler M., Regelsberger G., Jakopitsch C., Furtmuller P.G.,

RT Ruker F., Peschek G.A., Obinger C.;

RT "Nucleotide sequence analysis, overexpression in *Escherichia coli* and

RT kinetic characterization of *Anacystis nidulans* catalase-peroxidase.";

RL Biochimie 82:211-219(2000)

CC -|- CATALYTIC ACTIVITY: 2 H(2)O(2) -> O(2) + 2 H(2)O.

CC -|- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL

CC PEROXIDASE/CATALASE SUBFAMILY.

DR EMBL; AF197161; AAF05841.1; -.

DR HSSP; P00431; ICCK.

DR InterPro; IPR000763; Bac_ctase/prase.

DR InterPro; IPR002016; Peroxidase.

DR Pfam; PF00141; peroxidase; 1.

DR PRINTS; PR00458; PEROXIDASE.

DR TIGRFAMs; TIGR00198; cat_per_HPI; 1.

DR PROSITE; PS00436; PEROXIDASE 2; 1.

KW Heme; Hydrogen peroxidase; Iron; Oxidoreductase; Peroxidase.

SQ SEQUENCE 720 AA; 80119 MW; 196F15930700B463 CRC64;

Query Match 70.7%; Score 2870; DB 2; Length 720;

Best Local Similarity 72.1%; Pred. No. 4.9e-189;

Matches 520; Conservative 85; Mismatches 108; Indels 8; Gaps 3;

QY 35 AMSGKCPVMHGNTSTGTSTNKDWPEGLNLDLTHQDRKSDPMDPDFNYREVRKLDFDA 94

DB 3 ATRGKCPVMHGATTVNSTAEWPKALNLDLSQDRKTNPMGPDFNYQSEVKLDVAA 62

QY 95 LKKDVHALTSDQEWNPADWGHYGLMIRMAHWSAGTYRIADGRGGGTGQREAPLNSW 154

DB 63 LKQDLQALMTSDQDWNPADWGHYGLMIRLTHRAAGTYRIADGRGGAGTGNQREAPLNSW 122

QY 155 PNWSLDKARRLLWPIKKYGNKISWADLMILAGTVAYESMGLPAYGFSRQVDIWEPEK 214

DB 123 PDNTLNDKARRLLWPIKQYGNKLSWADLIAYAGTYIAYESMGLTFFGAFGREDIWHPEK 182

QY 215 DIYWGDEKWLAPS--DREY-GDVNKKPTMENPLAAVOMGLIYVNPGEVNGHPDPLRTA 270

DB 183 DIYWGPEKWPVPSNPNSRYTGD---RELENPLAAVMTGLIYVNPGEVNGDNPDLKTA 238

QY 271 QQVLETFARMANDEKTAALTAGGHTVGNCHGNASALAPDPKASDVENOGGLGNPNM 330

DB 239 HDVRVTFARMANDEETVALTAGGHTVGKCHGNAAALGPEPEGADVEDQGLGWINKTQ 298

QY 331 QGKASNAVTSGLGAWTNTPTFDMGYDPLLFGYVWELKKSPAGAHHEPDIKKNKPV 390

DB 299 SGIGRNAVTSGLGAWTPTPTQMDNGYFRMLNYDWELKKSPAGAWQEPINPREEDLP 358

QY 391 DASDPSIRNPIMTDADMAIKVNPYRAICEKFMADPEYFKTKFAKWFKLTHRDIGPKS 450

DB 359 DVEDPSIRNLVMTDADMAKKMDPEYRKISERYQDPAYFDVFAVAFWKLTHRDMDGPKA 418

QY 451 RYIGPEVPAEDLIWQDPIPAAGNTDYCEEVYKQIAQSGLSISEMYSTAWDSARTYRGSMD 510

DB 419 RYIGPDVPQEDLIWQDPIPAAGNRNDYQAVKDRIAAASGLSISELVSTAWDSARTYRNSDK 478

QY 511 RGGANGARTLAPQNEWQGNPERLAKVLSYEQIISADTGASIDVAVIAGSVGIEKAAK 570

DB 479 RGGANGARTLAPQKDWEGNEPDRILAKVLAVLEGTAATGASVADVIVLAGNVGVEQAAR 538

```

QY 457 VPAEDLLIWDPIPAAGNTDYCEEVVKQKIAQSGLSISEWSTANDSARTYRGSDMRGGANG 516
Db 429 VPAETLLIWDQVPPTGRKDYDVAKAKIIASGLSIGEMVSTANDSARTFRNSDKRGANG 488
QY 517 ARIRLAPQNEWQNEPERLAKVLSVYEQISADTGCASITADIVTLAGSVGIEKAAKAGYDV 576
Db 489 ARIRLAPQKDWLGNPEPERLAKVLSVYEQISADTGCASITADIVTLAGSVGIEKAAKAGIAI 548
QY 577 RVFPLKRGDATAEMTDADSFAPLEPLADGFRNMOKKKEYVYVVKPEMLLIDRAQLMGLGPE 636
Db 549 TVPFAAGRGDATIRQTDVSEFDVLEPIADGFRNMOKKQHYAVNPPELLIDRAQLMGLGSAPE 608
QY 637 MTVLLGCMRVLTNGYGTGKIGVFTDCGQLTNDFFVNLTDGMSWKKVPVGSNAYEIRDKT 696
Db 609 MTVLLIGLRLVLTNGHGTGKIGVFTDNGVGLSNDFFVNLTDGMSWKKVPVGSNAYEIRDKT 668
QY 697 GAVKWTASRVLDLVFGSNLSLRSVAEYVAQDDNGCKEKFVRDFVAAWTKVYNADRFDV 751
Db 669 GNVKWTATRVLDLVFGSNLSLRSVAEYVAQDDNGCKEKFVRDFVAAWTKVYNADRFDI 723

RESULT 5
Q80714 PRELIMINARY; PRT; 731 AA.
AC Q80714: 2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Catalase.
GN KATA OR ATU4642 OR AGR-L481.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP "The genome of the natural genetic engineer Agrobacterium tumefaciens C58."
RT Science 294:2317-2323(2001).
RN [2]
RP "Genome sequence of the plant pathogen and biotechnology agent Agrobacterium tumefaciens C58."
RT Science 294:2323-2328(2001).
DR EMBL; AE009392; AAL45436.1; ALT_INIT.
DR EMBL; AE008223; AAK88805.1; -.
DR InterPro; IPR000763; Bac_ctase/prase.
DR InterPro; IPR002016; Peroxidase.
DR InterPro; IPR01680; WD40.
DR Pfam; PF00141; peroxidase; 1.
DR PRINTS; PR00458; PEROXIDASE.
DR TIGRFAMs; TIGR00198; cat_per_HPI; 1.
DR PROSITE; PS00436; PEROXIDASE_2; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
KW Complete proteome.

```

```

SQ SEQUENCE 731 AA; 79570 MW; EIAA67F4E2456F20 CRC64;
Query Match 69.5%; Score 2820; DB 16; Length 731;
Best Local Similarity 71.08; Pred. No. 1.4e-185;
Matches 508; Conservative 86; Mismatches 121; Indels 0; Gaps 0;
QY 37 SGKCPVHGGNTSTGTSNKDWPEGLNLDILHOODRKSDDPDNFYEEYRKLDLDFALK 96
Db 16 AGKCPVHGGNTASGKSVTEWPNALNLDILHQHDTKTNPGLTSFNRYEALKTLDVEALK 75
QY 97 KDVEHALMTDSQEWHPADGHHYGGGLMIRMAHSHASCTYRIADRGCGGCTGSGORFAPLNSKPD 156
Db 76 ADRLALMTDSQEWHPADGHHYGGGLMIRMAHSHASCTYRIADRGCGGCTGSGORFAPLNSKPD 135
QY 157 NVSLDKARRLLWPIKKKYGNKISWADLMILAGTVAVESMGLPAYGFSFGRVDIWEPEKDI 216
Db 136 NVNTDKRRLLWPIKKKYGNKISWADLMILAGTVAVESMGLPAYGFSFGRVDIWEPEKDI 195
QY 217 YWGDEKEWLAPSDERYGDVANKPETMENPLAAVQMGIIYVNPGEVNGHPDPLRTAQOYLET 276
Db 196 YWGDEKEWLAPSDERYGDVANKPETMENPLAAVQMGIIYVNPGEVNGHPDPLRTAQOYLET 255
QY 277 FARMAMNDEKTAALTAGGHTVGNCHGNGNASALAPDPKASDVENQGLGHWGNPNMQGKASN 336
Db 256 FARMGMDDEETVALTAGGHTVGNCHGNGNASALAPDPKASDVENQGLGHWGNPNMQGKASN 315
QY 337 AVTSGIECAWTTNPKFDMGYFDLLFGYNWELKKSPAGAHHWEPIIDIKKNKPYVDASDPS 396
Db 316 TVVSGIEGAWTSEPTKWDNGFFDLFKHWTLTHSPAGASQWAPITTAEDKPYVDVEDAS 375
QY 397 IRHNPIMTDADMAIKVNPYTRAIACEKFMADPEYFKTKAKAWFKLTHRDLPKSRYSYRIGPE 456
Db 376 IRTIPNMTDADMAIKVNPYTRAIACEKFMADPEYFKTKAKAWFKLTHRDLPKSRYSYRIGPE 435
QY 457 VPAEDLLIWDPIPAAGNTDYCEEVVKQKIAQSGLSISEWSTANDSARTYRGSDMRGGANG 516
Db 436 VPAEDLLIWDPIPAAGNTDYCEEVVKQKIAQSGLSISEWSTANDSARTYRGSDMRGGANG 495
QY 517 ARIRLAPQNEWQNEPERLAKVLSVYEQISADTGCASITADIVTLAGSVGIEKAAKAGYDV 576
Db 496 ARIRLAPQKDWEGNEPARLSRVLSVLEPIARETGASITADIVTLAGSVGIEKAAKAGYDV 555
QY 577 RVFPLKRGDATAEMTDADSFAPLEPLADGFRNMOKKKEYVYVVKPEMLLIDRAQLMGLGPE 636
Db 556 AVPFAAGRGDASAEQTDADSFAPLEPLADGFRNMOKKKEYVYVVKPEMLLIDRAQLMGLGPE 615
QY 637 MTVLLGCMRVLTNGYGTGKIGVFTDCGQLTNDFFVNLTDGMSWKKVPVGSNAYEIRDKT 696
Db 616 LTVLLIGLRLVLTNGHGTGKIGVFTDNGVGLSNDFFVNLTDGMSWKKVPVGSNAYEIRDKT 675
QY 697 GAVKWTASRVLDLVFGSNLSLRSVAEYVAQDDNGCKEKFVRDFVAAWTKVYNADRFDV 751
Db 676 GAARYSATRVLDLVFGSNLSLRSVAEYVAQDDNGCKEKFVRDFVAAWTKVYNADRFDI 730

RESULT 6
Q80708 PRELIMINARY; PRT; 723 AA.
AC Q80708:
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Peroxidase/catalase (EC 1.11.1.6) (Catalase-peroxidase).
GN KATA.
OS Agrobacterium tumefaciens.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=358;
RN [1]
RP "SEQUENCE FROM N.A.
RT MEDLINE=20118000; PubMed=10652101;
RA Xu X.Q., Pan S.Q.;
RT "An Agrobacterium catalase is a virulence factor involved in

```

```

RT  tumorigenesis.*;
RL  Mol. Microbiol. 35:407-414(2000).
CC  -/- CATALYTIC ACTIVITY: 2 H(2)O(2) - O(2) + 2 H(2)O.
CC  -/- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL
CC  PEROXIDASE/CATALASE SUBFAMILY.
DR  EMBL: AB033631; BA089349.1; -.
DR  InterPro: IPR000763; Bac_ctase/prase.
DR  InterPro: IPR002016; Peroxidase.
DR  Pfam: PF00141; peroxidase; 1.
DR  PRINTS: PR00458; PEROXIDASE.
DR  TIGRFAMs: TIGR00198; cat_per_HPI; 1.
DR  PROSITE: PS00436; PEROXIDASE_2; 1.
DR  PROSITE: PS00678; WD_REPEATS_1; 1.
KW  Heme; Hydrogen peroxide; Iron; Oxidoreductase; Peroxidase.
SQ  SEQUENCE 723 AA; 78741 MW; AA531987F53B508 CRC64;

Query Match          69.4%; Score 2817; DB 2; Length 723;
Best Local Similarity 71.0%; Pred. No. 2.2e-185;
Matches 508; Conservative 85; Mismatches 122; Indels 0; Gaps 0;

QY  37  SGKCPVMHGNTSTGTSTNKNWPEGLNLDILHQQRKSDPMDPDFNYREVRKLDLFDALK 96
Db  :|||||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||
QY  8  AGKCPVMHGNTVSGRSVTEWPNALNLDLHQHDKTNPLGTSFNYREALKTLDYEALK 67
QY  97  KDVALMTDSQEWNPADWGHYGLMIRMAHWSAGTYRIADGRGGGTGSGQRFAPLNSWPD 156
Db  :|||||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||
QY  68  ADLRALMTDSQEWNPADWGSYVGMARVTHAAGSYRVTDGRGGANTGNQRFAPLNSWPD 127
QY  157  NVSLDKARRLLWPDKKYGKNSWADLMILAGTVAYESMGLPAYGFSGRVDIWEPEKDI 216
Db  :|||||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||
QY  128  NVNTDKRLLWPDKKYGKNSWADLIAGTIAYDVAGLTKFTGFAFGREDIWAPEKDT 187
QY  217  YNGDEKWLAPSDERYGDVKNKPTMENPLAAVOMGLIYVNPGEVNGHPDPLRTAQOVL 276
Db  :|||||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||
QY  188  YNGDEKWLAPSDGRYGDVSKPTLENPLAAVOMGLIYVNPGEVNGSKSPLATAAOMRET 247
QY  277  FARMANDEKTAALTAGGTVGCHNGNASALAPDPKASDVENQGLGNGNPNQKASN 336
Db  :|||||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||
QY  248  FARMGMDDEETVALTAGGTIGKSHGNSAANLSPDPAAGPYQGLGWINTKRGIGRD 307
QY  337  AVTSGIEGAWTNPFTFDMGYFOLLFGYNWELKKSAPGAHHWEPIDIKKENKPVDSAPS 396
Db  :|||||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||
QY  308  TVVSGIEGAWTSEPTKWDNGFFDMLPKHWTLTHSPAGASQWAPITIAEDKPVVDAS 367
QY  397  IRHNPINTDADMAIKNPTYRAICEKFMADPEYFKKTFKAWFKLTHRDIGPKSRYIGPE 456
Db  :|||||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||
QY  368  IRTIPMTDADMALKVDPIYREISLKFDDHFDSDVFARAWFKLTHRDWPKSRYVGP 427
QY  457  VPAEDLIWQDPIPAWNTDYCEEVVKOKIAQSGLSISEMVSATWSDARTYSGDMRGANG 516
Db  :|||||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||
QY  428  VPAEDLIWQDPIPAWNTDYCEEVVKOKIAQSGLSISEMVSATWSDARTYSGDMRGANG 487
QY  517  ARIRLAPQWQNEPERLAKVLSVYEQISADTGASTADIVVLGASVTEKAAGAAGYDV 576
Db  :|||||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||
QY  488  ARIRLAPQWQNEPERLAKVLSVYEQISADTGASTADIVVLGASVTEKAAGAAGYDV 547
QY  577  RVPFLAGRGDATAEMTDADSFAPLEADGFRWQKKEYVVKPEEMLLDRAQLGLTGP 636
Db  :|||||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||
QY  548  AVPFAAGRGDASAEQTDADSFAPLEADGFRWQKKEYVVKPEEMLLDRAQLGLTGP 607
QY  637  MTVLLGGMVGLNPGYGGTHGVPTDCEGQLTNDFFVNLTDGMSKWPVGSNAYEIRDKT 696
Db  :|||||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||
QY  608  LTVLIGGLRVIGANYGAAHGVPTDKPGALTDTFFTLTDMAYSWPTGNLLEIRDKT 667
QY  697  GAVKWTASRVLDVFGSNLLRSVAEYVAQDDNGEKEKFRDVFVAWTKVMNADRDV 751
Db  :|||||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||
QY  668  GAARYSATRVLDVIGSNLLRSVAEYVAQDDNGEKEKFRDVFVAWTKVMNADRDV 722

```

RESULT 7

Q8E981

ID Q8E981

PRELIMINARY;

PRT;

728 AA.

```

AC  Q8E981;
DT  01-MAR-2003 (TrEMBLrel. 23, Created)
DT  01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE  01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE  Catalase/peroxidase HPI.
GN  KATG-2 OR S04405.
OS  Shewanella oneidensis.
OC  Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC  Alteromonadaceae; Shewanella.
OX  NCBI_TaxID=70863;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=MR-1;
RX  MEDLINE=22297686; PubMed=12368813;
RA  Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA  Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA  Meyer T., Tsaplin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA  DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA  Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA  Vamathevan J., Weidman J., Imbraim L.A., Lee K., Berry K., Lee C.,
RA  Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA  Feildblum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.;
RT  *Genome sequence of the dissimilatory metal ion-reducing bacterium
RT  Shewanella oneidensis.*;
RL  Nat. Biotechnol. 20:1118-1123(2002).
DR  EMBL; A2015873; AAN57372.1; -.
DR  TIGR; S04405; -.
KW  Peroxidase; Complete proteome.
SQ  SEQUENCE 728 AA; 80980 MW; 354F1B25226EF0E2 CRC64;

Query Match          69.1%; Score 2802.5; DB 16; Length 728;
Best Local Similarity 71.1%; Pred. No. 2.2e-184;
Matches 513; Conservative 81; Mismatches 119; Indels 9; Gaps 4;

QY  38  GKCPVMHGNTSTGTSTNKNWPEGLNLDILHQQRKSDPMDPDFNYREVRKLDLFDALK 97
Db  :|||||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||
QY  9  GKCPVMHGNTSTGTSTNKNWPEGLNLDILHQQRKSDPMDPDFNYREVRKLDLFDALK 68
QY  98  DVHALMTDSQEWNPADWGHYGLMIRMAHWSAGTYRIADGRGGGTGSGQRFAPLNSWPD 157
Db  :|||||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||
QY  69  DLHSLMTDSQEWNPADWGHYGLMIRMAHWSAGTYRIADGRGGGTGSGQRFAPLNSWPD 128
QY  158  VSLDKARRLLWPDKKYGKNSWADLMILAGTVAYESMGLPAYGFSGRVDIWEPEKDI 217
Db  :|||||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||
QY  129  ANLDKARRLLWPDKKYGKNSWADLIAGTVAYESMGLTKFTGFAFGREDIWAPEKDI 188
QY  218  WGDEKWLAPSDERYGDVKNKPTMENPLAAVOMGLIYVNPGEVNGHPDPLRTAQOVL 277
Db  :|||||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||
QY  189  WGSEKWLAPSDPNRYSRGARDLENPLAAVOMGLIYVNPGEVNGHPDPLRTAQOVL 248
QY  278  ARMANDEKTAALTAGGTVGCHNGNASALAPDPKASDVENQGLGNGNPNQKASNA 337
Db  :|||||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||
QY  249  ARMANDEETVALTAGGTVGCHNGNAQDLGPEGEDEVAQGLGWLKKGSGVGCDA 308
QY  338  VTSIGIGAWTNPFTKWDNGFFDMLPKHWTLTHSPAGAHWEPIDIKKENKPVDSAPSI 397
Db  :|||||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||
QY  309  VTSLEGAWTTHPTQWDNGYFYLLNLYDWELKKSAPGAWQWEPINKEEDKVSVDPAK 368
QY  398  RHNPIMTDADMAIKNPTYRAICEKFMADPEYFKKTFKAWFKLTHRDIGPKSRYIGPE 457
Db  :|||||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||
QY  369  KFPNIMTDADMAIKIDPEYKRSERYQDPDPAYSEVFAFAWFKLTHRDIGPKSYLGPE 428
QY  458  PAEDLIWQDPIPAWNTDYCEEVVKOKIAQSGLSISEMVSATWSDARTYSGDMRG 512
Db  :|||||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||
QY  429  PSELLWQDPIPAWNTDYCEEVVKOKIAQSGLSISEMVSATWSDARTYSGDMRG 486
QY  513  GANGARIRLAPQWQNEPERLAKVLSVYEQISADTGASTADIVVLGASVTEKAAGA 570
Db  :|||||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||
QY  487  GANGARIRLAPQWQNEPERLAKVLSVYEQISADTGASTADIVVLGASVTEKAAGA 546
QY  571  AAGYDVVRVPLKRGDATAEMTDADSFAPLEADGFRWQKKEYVVKPEEMLLDRAQL 630
Db  :|||||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||
QY  547  AAGVNTTVPPMGRGDATKEMTDVESFVWLEPLHDAYRNWLKQDYVVVQPEEMLLDRS 606

```

```

QY 631 GLTGPEMTVLGGMRVLTGNYGGTKHGVTDCGGLTNDFFVNLDTMGNSMKRPGVSNAYE 690
DB 607 GLTAHEMTVLGGMRVLTGNYGGTKHGVTDRGVLSNDFVNLDTMAYSKRPVGTNLVE 666
QY 691 IDRKTKGAVKWTASRVLDVFGSNSLLRSYAEVYAQDDNGEKEKRVDFVAAWTKVMNADRPD 750
DB 667 VVERATGTVKWTATRVLDVFGSNSVLRVYAEVYAQDDAKEKRVDFVAAWTKVMNADRPD 726
QY 751 VA 752
DB 727 LA 728

RESULT 8
ID P73911 PRELIMINARY; PRT: 754 AA.
AC P73911; P97083;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Catalase HPI.
GN KATG OR SLI1987.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_taxid=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugitani M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT *Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.*;
RL DNA Res. 3:109-136(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Ushimaru T., Hayashi H., Murata N.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: D90910; BAA17975.1; -
DR EMBL: D83990; BAA20459.1; -
DR HSSP: P00431; IAEI.
DR InterPro: IPR000763; Bac_ctase/prase.
DR InterPro: IPR002016; Peroxidase.
DR Pfam: PF00141; peroxidase.1.
DR PRINTS: PR00458; PEROXIDASE.
DR TIGRFAMs: TIGR00198; cat_per_HPI; 1.
DR PROSITE: PS00436; PEROXIDASE_2; 1.
KW Complete proteome.
SQ SEQUENCE 754 AA; 84445 MW; AC5709D8B6468D4F7 CRC64;

```

```

Query Match 68.9%; Score 2795; DB 16; Length 754;
Best Local Similarity 69.8%; Pred. NO. 7.8e-184;
Matches 513; Conservative 79; Mismatches 131; Indels 12; Gaps 4;

QY 28 HRAIRERAM-----SGKCPVHGGTSTGTSGNKDWPEGLNLDLHQDRKSDPMDPDF 81
DB 18 HNRKERKPMANDQVPASKCPVHGGANTTGGQNLNWNPNALNLDLHQDRKTPMDPDF 77

QY 82 NYREVRKLPDALKVDHMLTDSQEWPPADWGHYGGGLMTRMAWHSAGTYRIADGRGG 141
DB 78 NYAEAFQQLDLAAVKQDLHLMLTDSQSWPPADWGHYGGGLMTRMAWHAAGTYRIADGRGA 137

QY 142 GTGQRAPLNSWPDNVSLDKARLLWPKKKYGNKISWADMLIAGTVAYESMGLPAYG 201
DB 138 ATGNRFAPLNSWPDNVSLDKARLLWPKKKYGNKISWADMLIAGTVAYESMGLPAYG 197

QY 202 FSFGKVDIWEKDIYGEKELAPSDERYGDYKPTMENPLAAVOMGLIYVPEGVN 261
DB 198 FAGGREDIWHPEKDIYGAKEWLASDHRIGSEDR-ESLENPLAAVOMGLIYVPEGVN 256

```

```

QY 262 GHPPDLRTAQVLETFARMAMNDEKTAALTAGGHTVGNCHGNASALAPDPKASDVNQ 321
DB 257 GHPPDLCTAQDVRTTFARMAMNDEETVALTAGHTVGNCHGNKAEILPEPEGADVQEQ 316
QY 322 GLGWNPNQKSNAGNAVTSIGEGAWTNTPTKEDMGYFDLLFGYNMELKKSPAGAHHEPI 381
DB 317 GLGWNQNGKGVGRETMSGIEGAWTTHPTQWDNGYFYMLFNHEWELKKSPAGAWQWEPV 376
QY 382 DIKKENKPVDAADSPSIRHNPIMTDADMAIKVNPYRAICEKFNADPEYFKKFAKAWFKL 441
DB 377 NIKEEDKPVDEDPNIRHNPIMTDADMAIKDPIYRQISERYREPDYFAEYFAKAWFKL 436
QY 442 THRDLPKRSYIGPEVPAEDLIWQDPIPAAGNTDYCEEVVK---QKIAQSGLSISEMVSTA 498
DB 437 THRDLPKRSYLGPDVQEDLIWQDPIPDVDTLSEGEKELEQOILLASGLIVSELVCTA 496
QY 499 WDSARTYRGSMDRGGANGARIRLAPQNEWQGNPEPERLAKVLSVYFOISADTG--ASIAADV 556
DB 497 WDSARTFSSDYRGGANGARIRLEPKNPNWGNPTRLAKVLAVLENIQANFAKPVSLADL 556
QY 557 IVLAGSVGIEKAAAGYDVRVPFLKGRGDATADSDADSPLEPLADGFERNWOKKEVY 616
DB 557 IVLGGGAAAKAALDGGIEVNVFPLPGRGDATQAMTDAESFTPLEPIHDGYRNWLKQDYA 616
QY 617 VKPEEMLLDRAQLMGLTGPMTVLGGMRVLTGNTGGTKHGVTDCGGLTNDFFVNLDT 676
DB 617 VSPPELLERTQLMGLTAPMTVLGGMRVLTGNTGGTKHGVTDRGVLSNDFVNLDT 676
QY 677 MGNWSKPVGNSNAYEIRDRKTKGAVKWTASRVLDVFGSNSLLRSYAEVYAQDDNGEKEVDF 736
DB 677 MAYQWRPAGNLLYEIGDQRTGEVKTATKVDLVFGSNSILRSYAEVYAQDDNREKRVDF 736
QY 737 VAAWTKVMNADREFV 751
DB 737 VAAWTKVMNADREFDL 751

```

RESULT 9

```

Q55110 PRELIMINARY; PRT: 720 AA.
AC Q55110;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Catalase-peroxidase (EC 1.11.1.6) (Peroxidase/catalase).
OS Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_taxid=1140;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC7942;
RX MEDLINE=96235199; PubMed=8645214;
RA Matsuda M., Ishikawa T., Takeda T., Shigeoka S.;
RT *The catalase-peroxidase of synechococcus PCC 7942: Purification,
RT nucleotide sequence analysis and expression in escherichia.coli.*;
RL Biochem. J. 316:251-257(1996).
CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) -> O(2) + 2 H(2)O.
CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL
CC PEROXIDASE/CATALASE SUBFAMILY.
DR EMBL: D61378; BAA09601.1; -
DR HSSP: P00431; 2CYP.
DR InterPro: IPR000763; Bac_ctase/prase.
DR InterPro: IPR002016; Peroxidase.
DR Pfam: PF00141; peroxidase; 1.
DR PRINTS: PR00458; PEROXIDASE.
DR TIGRFAMs: TIGR00198; cat_per_HPI; 1.
DR PROSITE: PS00436; PEROXIDASE_2; 1.
KW Heme; Hydrogen peroxidase; Iron; Oxidoreductase; Peroxidase.
SQ SEQUENCE 720 AA; 79910 MW; FA5053CF7C47C513 CRC64;

```

Query Match 67.3%; Score 2731; DB 2; Length 720;
Best Local Similarity 69.8%; Pred. NO. 1.9e-179;

Matches	504;	Conservative	83;	Mismatches	125;	Indels	10;	Gaps	5
Qy	35	ΔMSGKCPVMHGNGTSTGT'SNKKDWMEGLNLDILHQDQKSDPMDPDPDFNYREYVRKLD	94	1	1	1	1	1	1
Db	3	ATQCKCPVMHGGAATTNLSIAEWWPKALNLDLSQHDRKTNPMGDPDFNYQEYVKLD-AA	61	1	1	1	1	1	1
Qy	95	LKDDVHALMTDSQEWPAADGHYGGILIRMAWHSAGTYRIADGRGGGGTGSQRFAPLNSW	154	1	1	1	1	1	1
Db	62	LKQDLQALMTDSQDWPADGHYGGILIRLTHWAAGTYRIADGRGGAGTGQRFAPLNSW	121	1	1	1	1	1	1
Qy	155	PDNYSLDKARLLWPIPKKYGKNTKSADWMLTACTVAYESMGLPAYGSGFRGVDIWEPEK	214	1	1	1	1	1	1
Db	122	PDNTNLDKARLLWPIKQYKGNKLSWADLIAAYGTIATYESMGLTFFGAFGREDLWHPEK	181	1	1	1	1	1	1
Qy	215	DIYWGDEKEWLAPS--DERY-GDVKRPETMENPLAAVQMGLIYVNPESGVNGHPDLRTA	270	1	1	1	1	1	1
Db	182	DIYWGPEKEWFPPTNPNSRYTGD---RELENPLAAVTMGLIYVNPESGVDGNPDPLKTA	237	1	1	1	1	1	1
Qy	271	QOVLLETARMAANDEKTAALTAGGTVGNGHNGNASALADPDPKASDVENGOLGWNPM	330	1	1	1	1	1	1
Db	238	HDVRVTFARMAANDEETVALTAGGTVGKCHGNGNAALLGPEPEGADVEDOGLGWINKTQ	297	1	1	1	1	1	1
Qy	331	QKQASNAVTSIGIEGAWTNTPTKFDWGFDDL-FGYNNELKESPGAHHWEPIDITKKENK	389	1	1	1	1	1	1
Db	298	SGIGRNAVTSGLEGAWTPHTQWDONGYFVCSLNTDWELKKNPAGAWQWEPINPREEDLP	357	1	1	1	1	1	1
Qy	390	VDASDPSIRHNPIMTDADMAIKVNPTYRAICEKFMADPEYFKKTFAKAWFKLTHRDLGPK	449	1	1	1	1	1	1
Db	358	VDVEDPSIRRLNVMTDADMAKMDPEYRKISERFYQDPAYFADYFARAWFKLTHRDGMPK	417	1	1	1	1	1	1
Qy	450	SRYTGPEVPAEDLIWQDPIPAAGNTDYCEEVVKQIAOGLSISEMVTASDARTYRGS	509	1	1	1	1	1	1
Db	418	ARYTGPDPVQDLWQDPIPAAGNPNYDYQAVKDRIAASGLSISELVSTASDARTYRNS	477	1	1	1	1	1	1
Qy	510	MRGANGARIRLAPONEWQNEPERLAKVLSVVEQISADTCASADTVTLVAGSVGIERAA	569	1	1	1	1	1	1
Db	478	KRGANGARIRLAPQKWEGNEPDRLPKVLAVLEGISAATGATVADVTLVAGNVGVEOKA	537	1	1	1	1	1	1
Qy	570	KAAGYDVRVPEPLKRGDATAEMTDADSFAPLEPLADGFRNKKQKEYVYVKPEMLIDRAQL	629	1	1	1	1	1	1
Db	538	RAAGVEIVLPAPGRGDATATQDTESEFAVLEPHDATATGSSRTMRQRLKNCCLIAIATQL	597	1	1	1	1	1	1
Qy	630	MGLTGPENTVLLGMRVLGTYNGGTKKGVTDFDCEQLTNDFFVNLITDMGNSWKPVGSNAY	689	1	1	1	1	1	1
Db	598	LGLTAPETVLLIGLRVLGTNHGGTKHVVFTDREGVLTNDFVNLITDMNLYLWKPKAGKNLY	657	1	1	1	1	1	1
Qy	690	EIDRRKTCVAKWTASRDVLTVGSSNLSLSYAEVTAQDNGEKKFYRDFVAATKVMNADRF	749	1	1	1	1	1	1
Db	658	EICDRKTNQVAKWTATRDVLTVFGSSNLSILRAYSELTAQDDNKEKFFVRDFVAATKVMNADRF	717	1	1	1	1	1	1
Qy	750	DV 751							
Db	718	DL 719							
RESULT 10									
Q9KEE6	ID	Q9KEE6	PRELIMINARY;	PRG:	735	AA.			
AC	Q9KEE6:								
DT	01-OCT-2000	(TREMBLrel. 15, Created)							
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)							
DT	01-OCT-2002	(TREMBLrel. 22, Last annotation update)							
DE	Catalase.								
GN	BH0906								
OS	Bacillus halodurans.								
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.								
OX	NCBI_TaxID=86665;								
ON	[1]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN-C-125 / JCM 9153;								
RX	MEDLINE=20512582; PubMed=11058132;								
RA	Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,								
RA	Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kubara S.,								

RA	Horikoshi K.;		
RT	*Complete genome sequence of the alkaliphilic bacterium Bacillus		
RL	halodurans and genomic sequence comparison with Bacillus subtilis.*;		
DR	Nucleic Acids Res. 28:4317-4331(2000).		
DR	EMBL; AP001510; BAB04625.1; -		
DR	HSSP; P00431; 1A2F.		
DR	InterPro; IPR000763; Bac_ctase/prase.		
DR	InterPro; IPR002016; Peroxidase.		
DR	Pfam; PF00141; peroxidase; 1.		
DR	PRINTS; PR00458; PEROXIDASE.		
DR	TIGRFAMS; TIGR00198; cat_per HPI; 1.		
DR	PROSITE; PS00435; PEROXIDASE_1; 1.		
DR	PROSITE; PS00436; PEROXIDASE_2; 1.		
DR	Complete proteome.		
SW	SEQUENCE 735 AA; 81530 MW; E93B4B5EBFA3A593 CRC64;		
Query Match	62.0%;	Score 2514;	DB 16; Length 735;
Best Local Similarity	64.1%;	Pred. No. 1.7e-164;	
Matches	464;	Conservative 90;	Mismatches 156; Indels 14; Gaps 6;
Qy	37	SGKCPVWHGGNT---STGTSNKDWPPEGNLNLIILHQODRKSDPMDPDFNYEEVKRLD	93
Db	6	TGKCPFHGGATTSPKSSGTTNDWPNALNLIHQDKSNPMGEEFDYAEFSKLDYD	65
Qy	94	ALKKDYHALMTDSQEWNPADWGHYGGIMTMAWHASAGTVRIADRGGGTGSORFAPLNS	153
Db	66	ALAQDYRLMRSDQWNPADFGHYGPFRTMSWHAAGTVRIADRGGGGTGNQRFAPLNS	125
Qy	154	WPNVSLDKARRLLWPIKKYGNKISWADMLAGTVAYESMGLPAYGFSGRVDIWEPE	213
Db	126	WPONGNLDKARRLLWPIKQYGNKISWADLLVLAGNVATEDMGGVPVIGFAGREDIWIPE	185
Qy	214	KDIYWGDEKWLAPSDERY-GDVNKPETWENPLAAVQGLIYVNPGEVNGHPDPLRTAQQ	272
Db	186	EDIYWGSEKWL-T-GDKRSGD----RELENPLAAVEMGLIYVNPGEVNGHPDPLRTAAHD	240
Qy	273	VLETFARMANDEKTAALTAGGHTVGNCHGNASALAPDPKASDVENQGLGWGNPMQG	332
Db	241	IRETFGRMGNDSETVALLAGGHTFGKAHGAGNPDHVGPPEAPAEIQAQGLGWQNTYSG	300
Qy	333	KASNATVSGEGAWTNPPTKFMQYDILFGYNWELKSPACAHWIEDIDIKENKPYDA	392
Db	301	KGNDTITSGLEGAWTANPTQWDNGFPDLFGYEMWLTKSPAGAYQWQAVDPDEKDLAPDA	360
Qy	393	SDPSIRHNPIMTDADMAIKVNPITYRAICEKFMADPEYFKKTFKAKFWKLTIRHDLGPKSRY	452
Db	361	EDFSVKVPTVMTLTDLARHDPPEYKISRREKHNPFADAPARAWFKLLHEDMGPKARY	420
Qy	453	IGPEVPAEDLIWQDPTPAGN---TDYCEEVVKQKTAQSGLSITSEMVSTAWQASARTYRGS	509
Db	421	LGPPEVPAEDFIWQDPTPYDYELTDAAVEVELKAKLIDSLGTVSELVYSELVYSELVYSELV	480
Qy	510	MRRGANGARIRLAPQEWQGNENERLAKVLSVVEQISA--DFGASADVIVLAGSVGTEK	567
Db	481	KRGANGARIRLAPQKDWQVNPQLEKVLVLENLQSQLDKKVSIALDLVLGSSAAVEK	540
Qy	568	AKAAGYDVVRVPLKGRGDATAMTDADSFAPLEPLADGFRNWKQKEYVYRPEEMLLDRA	627
Db	541	AAKEAGFDVTVFAPGRGDATQBDTVEGFVAVLEPVSQDGRNFKYQKKEYSVSPSEELLIDKA	600
Qy	628	QLMGLTGPEMTVLLGMRVILGTYNGGTKHGVTDCGQLTNDFFVNLITDMGNSWKPVGSN	687
Db	601	QLDLTLAPENTALIGMRALGANYGQTQHGVTDCVGTLTNDFFVNLITDMGIEWKRPVDYN	660
Qy	688	AYETIRKTKGAVKWTASRVLDLVFGNSLSLRSAEYVAAQDDNCEKFFVRDFVAAWTKVMNAD	747
Db	661	LYEGDRKTKGEVYRTATRVLDLVFGNSILIRAEYVAAQDDNCKGFVEDFIAAWKVMNAD	720
Qy	748	RFVD 751	
Db	721	RFDL 724	


```

Db      237 AAADIRETFRMGNDEETVALLTAGHTFGKAGAGPATHVGPPEAPAEIAGLWIS 296
Qy      328 PNMCKASNAVTSIGEGAWTPTKFDGMYFDLLFGYNWELKSPAGAHHEPIDIKEN 387
Db      297 SYGKSGSDTITSGIEGAWTPTQWDTSYDMLFGYDMLTKSPAGAWQMAVDPDEKD 356
Qy      388 KPVASDPSIRHNPIMTDADMAIKVNTYRAICEKFMADPEYFKKTKAKWFKLTHRDLG 447
Db      357 LAPDAEDPSKKVPTMMMTDLALRFDPEYKIAIRFHONPEEFAEAFARAWFKLTHRDMG 416
Qy      448 PKSYIGPEVPAEDLIWQDPIAGNTDYCE---EWWKOKIAQSGLSISEMYSTAWDSART 504
Db      417 PKTYLGLPEVPKEDFIHQDPIPEVDYELTEAEIEIEIKAKILNSGLTVSELVKTAWASAST 476
Qy      505 YRSDMRGGANGARIRLAPQNEWGNPEPERLAKVLSVYEQISAD--TCASIADVIVLAGS 562
Db      477 FRNSDKRGANGARIRLAPQKDWENPEPERLAKVLSVYEDIQRELPKKVSIALDLVLGGS 536
Qy      563 VGIEKAAAGYDVRVFLKRGDATAEMTDADSFAPLEPLADGFRNWKKEYVYVPEEM 622
Db      537 AAVEKAARDAGFDVKVFPFGRGDQTOQDTSVESFAVLEPPADGFRNWKKEYVYVPEEL 596
Qy      623 LLDRALQMLGTPENTVLGGMVRLGTYNGGTHGVTDCBGLTNDFFVNLTDGNSWK 682
Db      597 LVDRALQMLGTPENTVLGGLRVLGANYRDLPHGVFTDRIGVLTNDFFVNLDMNTEYV 656
Qy      683 PVGSNAYEIRDRKTKGAVKWTASRVDLVFGSNLSLSYAEVYAQDNGEKFVRDFVAATWK 742
Db      657 PTDSGIYEIRDRKTKGEVWTRATRVLDLIFGNSILRSYAEFFAQQDNGEKFVRDFINAWK 716
Qy      743 VNNADRFV 751
Db      717 VNNADRFDL 725

```

RESULT 13

```

Q955N2 ID Q955N2 PRELIMINARY; PRT; 744 AA.
AC Q955N2;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Peroxidase/catalase (EC 1.11.1.6) (Catalase-peroxidase).
GN CAT.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RA Matsuura T., Miyai K., Trakulnaleamsai S., Yomo T., Shima Y.,
RA Miki S., Yamamoto K., Urabe I.;
RT "Evolutionary molecular engineering by random elongation
mutagenesis."
RL Nat. Biotechnol. 17:58-61(1998).
CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL
CC PEROXIDASE/CATALASE SUBFAMILY.
DR EMBL; AB020092; BAA37004.1; -.
DR HSSP; P00431; 1B39.
DR InterPro; IPR000763; Bac_catase/prase.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00141; peroxidase; 1.
DR PRINTS; PR00458; PEROXIDASE.
DR TIGRFAMs; TIGR00198; cat_per_HPI; 1.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS00436; PEROXIDASE_2; 1.
KW Heme; Hydrogen peroxide; Iron; Oxidoreductase; Peroxidase.
SQ SEQUENCE 744 AA; 83946 MW; 34632D2BB5B109BB CRC64;

```

Query Match

Best Local Similarity 61.3%; Score 2489; DB 2; Length 744;

Matches 468; Conservative 85; Mismatches 160; Indels 16; Gaps 7;

```

Qy      32 RERAMSKCPVHGGNT---STGTSNKDWPEGLNLDILHODRKSDDPDPDFNYREVR 88
Db      4 QNRQNAACP-FHGSVTNQSSNRTNKDWPNQLNLSILHQDRKTNPHDEFEYAEFFQ 62
Qy      89 KLDFDALKKVYHAKWTSQEWKPADWGHYGLIRMAHWSAGTYRIADGRGGTGSORF 148
Db      63 KLDYALKEDURKLTESQDWWPADYGHYGLIRMAHWSAGTYRIADGRGGTGSORF 122
Qy      149 APLNSWPDNVSLDKARLLWPKKKYGNKISWADLMILAGTVAYESMGLPAYGFSFGVSD 208
Db      123 APLNSWPDNANLDRARLLWPKKKYGNKISWADLFILAGNVATSMGKKTIGFSGGRVD 182
Qy      209 IWEPEKDIYGDKEWELAPSERY-GDVNKKETWENPLAAVQMGGLIYVNPYGVNHPDPL 267
Db      183 VVHPEEDVYWGSEKWL- -SERYSGD- - -RELENPLAAVQMGGLIYVNPYGVNHPDPL 236
Qy      268 RTAOVLETFARMANDEKTAALTAGGHTVNCNGHGNASALAPDPKASDVENQGLWGN 327
Db      237 AAADIRETFRMGNDEETVALLTAGHTFGKAGAGPATHVGPPEAPAEIAGLWIS 296
Qy      328 PNMCKASNAVTSIGEGAWTPTKFDGMYFDLLFGYNWELKSPAGAHHEPIDIKEN 387
Db      297 SYGKSGSDTITSGIEGAWTPTQWDTSYDMLFGYDMLTKSPAGAWQMAVDPDEKD 356
Qy      388 KPVASDPSIRHNPIMTDADMAIKVNTYRAICEKFMADPEYFKKTKAKWFKLTHRDLG 447
Db      357 LAPDAEDPSKKVPTMMMTDLALRFDPEYKIAIRFHONPEEFAEAFARAWFKLTHRDMG 416
Qy      448 PKSYIGPEVPAEDLIWQDPIAGNTDYCE---EWWKOKIAQSGLSISEMYSTAWDSART 504
Db      417 PKTYLGLPEVPKEDFIHQDPIPEVDYELTEAEIEIEIKAKILNSGLTVSELVKTAWASAST 476
Qy      505 YRSDMRGGANGARIRLAPQNEWGNPEPERLAKVLSVYEQISAD--TCASIADVIVLAGS 562
Db      477 FRNSDKRGANGARIRLAPQKDWENPEPERLAKVLSVYEDIQRELPKKVSIALDLVLGGS 536
Qy      563 VGIEKAAAGYDVRVFLKRGDATAEMTDADSFAPLEPLADGFRNWKKEYVYVPEEM 622
Db      537 AAVEKAARDAGFDVKVFPFGRGDQTOQDTSVESFAVLEPPADGFRNWKKEYVYVPEEL 596
Qy      623 LLDRALQMLGTPENTVLGGMVRLGTYNGGTHGVTDCBGLTNDFFVNLTDGNSWK 682
Db      597 LVDRALQMLGTPENTVLGGLRVLGANYRDLPHGVFTDRIGVLTNDFFVNLDMNTEYV 656
Qy      683 PVGSNAYEIRDRKTKGAVKWTASRVDLVFGSNLSLSYAEVYAQDNGEKFVRDFVAATWK 742
Db      657 PTDSGIYEIRDRKTKGEVWTRATRVLDLIFGNSILRSYAEFFAQQDNGEKFVRDFINAWK 716
Qy      743 VNNADRFV 751
Db      717 VNNADRFDL 725

```

RESULT 14

```

Q955K7 ID Q955K7 PRELIMINARY; PRT; 744 AA.
AC Q955K7;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Peroxidase/catalase (EC 1.11.1.6) (Catalase-peroxidase).
GN CAT.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RA Matsuura T., Miyai K., Trakulnaleamsai S., Yomo T., Shima Y.,
RA Miki S., Yamamoto K., Urabe I.;
RT "Evolutionary molecular engineering by random elongation
mutagenesis."
RL Nat. Biotechnol. 17:58-61(1998).

```

CC	-I- CATALYTIC ACTIVITY: 2 H(2)O(2) - O(2) + 2 H(2)O.
CC	-I- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL
CC	PEROXIDASE/CATALASE SUBFAMILY.
DR	EMBL; AB020067; BAA36979.1; --
DR	HSSP; P00431; 1B79.
DR	InterPro; IPR000763; Bac_ctase/prase.
DR	InterPro; IPR002016; Peroxidase.
DR	Pfam; PF00141; peroxidase; 1.
DR	PRINTS; PR00458; PEROXIDASE.
DR	TIGRFAMs; TIGR00198; cat_per_HPI; 1.
DR	PROSITE; PS00435; PEROXIDASE_1; 1.
DR	PROSITE; PS00436; PEROXIDASE_2; 1.
KW	Heme; Hydrogen peroxide; Iron; Oxidoreductase; Peroxidase.
SQ	SEQUENCE 744 AA; 61.00 MW; 35CEED2142D4F6AB CRC64;
	Query Match
	Best Local Similarity 64.2%; Score 2489; DB 2; Length 744;
	Matches 468; Conservative 85; Mismatches 160; Indels 16; Gaps 7;
Qy	32 RERAMSKCPVHNGGNT---STGSNKDWMPGLNLIDLHQDDRKSDPMDFNYREVR 88
Dd	4 QNRQAACQP-FHGSVTNQSSNRTNKDWPNQLNSLTHQHDKRNPHDEEFYAEBFQ 62
Qy	89 KLDPDAKKOVHALMTDSOEWWPADGHYGGLTIRMAHWSAGTYRIADRGCGGTGSGRF 148
Dd	63 KLDYIAWKEDURKLWTESQWPPADYGHYGPLFIIRMAHWSAGTYRIGDGRGAESTGTQRF 122
Qy	149 APLNSWPDNLSLDRARLLMPIKKKYGNKISWADMLITAGTVAYESMGLPAYGFSFRVD 208
Dd	123 APLNSWPDNANLDRARLLMPIKKKYGNKISWADLFILAGNVATESMGKTIGFGGRVD 182
Qy	209 IWPEKDIYWGDEREWLAPSDERY-GDVNRPETMENPLAIVOMGLLIYNPEGVNHDPDL 267
Dd	183 VMHPDEDVYGGSEKWLA--SERYSGD----RELLENLAAYQMGLLIYNPEGDGCKPDPK 236
Qy	268 RTAOVLLETFAARMNDKTAALTAGHTVCNCNGHGNASALAPDKASDVENOGLGWN 327
Dd	237 AAARDIETFRMGNDDEETVALLAGHTTGKAHGAGPTHVGPPEPAAPTEAGLGWIS 296
Qy	328 PNMOGKASNVAVTSGIEGAWTNPTKFDWGYFDLLFGYNWELKKSPAGAHHWEPIDIKKN 387
Dd	297 SYGKGSDTTITSGIEGAWTPPTQWDTSYFDMLFGYDWMLTKSPAGAQWMAVDPDEKD 356
Qy	388 KPVDASDPSIRHNPIMTDAMIAKVNTYTRAICEKFMADPEYFKTKAKWFKLTHRDLG 447
Dd	357 LAPDAEEDSKVPYMMMTDLALRFDPEYEKIARFRHQNPEEFAEAFARAWFKLTHRDVG 416
Qy	448 PKSYIGPEVPAEDLIWODPIPAGNTDYCE---EWVKOKIAQSGLSISEMSTAWSART 504
Dd	417 PKTRYLGPVEPKEDIWQDPIPEVDYELTEAEIEEKAKILNSGLTVSELVKTAWASAST 476
Qy	505 YRGSDMRGGANGARIRLAPONEWOGNEPERLAKVLSVYEQISAD--TCASADIYVILAGS 562
Dd	477 FRNSDKRGGANGARIRLAPQDWVNENEPEARLAKVLSVYEDIQRELPKKVSIAIDLVLGGS 536
Qy	563 VGITEKAKAAGYDVRVFLKRGCDATAMTDADFAPLEPLADGFRRWQKEYYVVKPEEM 622
Dd	537 AAVEAKAARDAGFDVKVPFFPGRGDATQEOTDVFESFAVLEPFADGFRNYQKQYVPPPEL 596
Qy	623 LLDRALQMLGTLGPENTVLLGMRYLGNFYGGTKGVFTDCEGOLNTDFFVNLTMDSNWK 682
Dd	597 LVDKAQLLGLTAPENTVLVGLRVLGANYRDLPHGVFTDRIGVUTNDDFFVNLDMNTIEWV 656
Qy	683 PVGSNAYEIRDRKTKGAVKWTAASRVLDLVFGSNLSRSYAEEYAQDDNGEKFVRDFVAAVTK 742
Dd	657 PTDSGIYEIRDRKTGEYRWATRYLDLVFGSNLSRSYAEEYAQDDNQEKFFVRDFINAWVK 716
Qy	743 VMNAADRDFV 751
Dd	717 VMNAADREFL 725

RESULT 15

Q9S5Q5	PRELIMINARY;	PRT;	745 AA.
AC	Q9S5Q5		
DT	01-MAY-2000 (Tremblrel. 13, Created)		
DT	01-MAY-2000 (Tremblrel. 13, Last sequence update)		
DT	01-MAR-2003 (Tremblrel. 23, Last annotation update)		
DE	Peroxidase/catalase (EC 1.11.1.6) (Catalase-peroxidase).		
GN	CAT.		
OS	Bacillus stearothermophilus.		
OC	Bacteria; Firmicutes; Bacillales; Geobacillus.		
OX	NCBI_TaxID-1422;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Matsuura T., Miyai K., Trakulnaleaamsai S., Yomo T., Shima Y.,		
RA	Miki S., Yamamoto K., Urabe I.;		
RT	*Evolutionary molecular engineering by random elongation		
RT	mutagenesis*;		
RL	Nat. Biotechnol. 17:58-61(1998).		
CC	-1- CATALYTIC ACTIVITY: 2 H(2)O(2) -> O(2) + 2 H(2)O.		
CC	-1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL		
CC	PEROXIDASE/CATALASE SUBFAMILY.		
DR	EMBL; AB020115; BAA37027.1; -.		
DR	HSSP; P00431; 1839.		
DR	InterPro; IPR000763; Bac_ctase/prase.		
DR	InterPro; IPR002016; Peroxidase.		
DR	Pfam; PF00141; peroxidase; 1.		
DR	PRINTS; PR00458; PEROXIDASE.		
DR	TIGRFAMS; TIGR00198; cat_per_HPI; 1.		
DR	PROSITE; PS00435; PEROXIDASE_1; 1.		
DR	PROSITE; PS00436; PEROXIDASE_2; 1.		
KW	Heme: Hydrogen peroxide; Iron; Oxidoreductase; Peroxidase.		
SQ	SEQUENCE 745 AA; 84135 MW; 92656ED88814B5272 CRC64;		
Query Match 61.3%; Score 2489; DB 2; Length 745;			
Best Local Similarity 64.2%; pred. No. 9.1e-163;			
Matches 468; Conservative 85; Mismatches 160; Indels 16; Gaps 7;			
Qy	32	REARMSKCPVHGNGT---STGTSKNDWPPEGLNLDILHQDRKSDPDPDPFNREEVR	88
Db	4	QNRQNAACP-FHGSVTNQSNRTNKDWPQNQLNSILHQDRKTNPHDEEFNAAEFQ	62
Qy	89	KLDPALKDKVHALMTDSQSWPMDWGHYGLMIRMAWHAGTYRIADRGGGTGSORF	148
Db	63	KLDYALKEDLRKLMTESQDWPADYGHYGLFIRMAWHAGTYRIADRGGGTGTQRF	122
Qy	149	APLNSWPDNVLDKARLLPIPKKKYGNKTSWADLMTACTVAYESMGLPAYGFSGRVD	208
Db	123	APLNSWPDNAMLKARLLPIPKKKYGNKTSWADLFTLAGNVAIESMGKTYIGFGGRVD	182
Qy	209	IWEPEKDIYGDKEWLAPSDERY-GDVKPMTENPLAAVQGLIYVNPGEVNGHPDPL	267
Db	183	VWHEEDVYWGSEKWL--SERYSGD---RELENPLAAVQGLIYVNPGEVNGHPDPK	236
Qy	268	RTAQVLETFARMNDKTAALTAGGTGYNGHGNASALAPDPKASDVENQGLGWN	327
Db	237	AAARDIRETFRMGMNDTEVALLAGGHTGKAHGAGATHVGPPEAAPIEAQGLGIS	296
Qy	328	PNWOCKASNAVTSIGCAWNTNPKFDMGYFDLLFGYNWELKKSPAGAHWEIDIDTKEN	387
Db	297	SYGKSGSDITTSIGCAWNTPTQMDTSTYFDMLFGYDMWLTSPAGAWQMAVDPEKD	356
Qy	388	KPVDASPSRTHNPIMTDADMAIKVNPYRAICEKFMADPEYKPKTKAKAWFKLTHRDLG	447
Db	357	LAPDAEPSKRVPTMMVTDLALRFDEYEKIARRFHQNPDEEAFARAWFKLTHRDWG	416
Qy	448	PKSYIGPEVPAEDLIWQDPIPAAGNTDYCE---EVVKQKTAQSGLTISEKSVTAWDSART	504
Db	417	PKTRYLGPVPEKDFIWQDPIPEVDVYELTAEETEEIKAKILNSGLTVSELVKTKAWASAST	476
Qy	505	YRGSDMGGGANGARILRAPONEWQGNPEKALVLSVYVEISAD--TGASIADVIVLAGS	562
Db	477	FRNSDKRGKGANGARILRAPQKDEKNEVPEKALVLSVYEDIQRELPKVKVIADLIVLGGS	536

QY	563	VGIEKAAAGYDVRVPEFLKRGDATAEMTDADSFAPLEPLADGFRNWKKEYVVKPEEM	622
Db	537	AAVEKAAARDAGFDVKVPFPFGRGDATQEQTDVESFAVLEPFADGFRNYQKQYSVPPEEL	596
QY	623	LLDRAQLMGLTGPEMTVLIGGNRVLTGTYGGTKHGVTDCGOLTNDFEVNLTDMGNSWK	682
Db	597	LYDKAQLLGLTAPEMTVLVGGLRVLGANYRDLPHGVFTDRIGVLTNDFEVNLLDMYEWV	656
QY	683	PVGSNAYEIRDRKTKGAVKWTASRDVLVFGSNLLRSYAEVYAQDDNGEKFVRDFVAATK	742
Db	657	PTDSGIYEIRDRKTGEVRWTATRVDLIFGSNSILRSYAEFYAQQDDNQEFVRDFINAWVK	716
QY	743	VNADREFV	751
Db	717	VNADREFDL	725

Search completed: October 7, 2003, 19:45:39
 Job time : 81.4005 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2003, 06:25:58 ; Search time 8119.46 Seconds
(without alignments)
11276.082 Million cell updates/sec

Title: US-09-884-889-7
Perfect score: 2238
Sequence: 1 atggaaatcacaacactc.....accggttgatctgaataa 2238

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_in.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2238	100.0	2238	6	AR098264	AR098264 Sequence
2	2238	100.0	2238	6	AR215252	AR215252 Sequence
3	1053.2	47.1	11320	1	AE010762	AE010762 Methanosa
c	1024.2	45.8	198050	1	AL646061	AL646061 Ralstonia
5	991.8	44.3	340857	1	AP003010	AP003010 Mesorhizo
c	960	42.5	10578	1	AE007314	AE007314 Sinorhizo
6	951.2	42.5	311963	1	AE016872	AE016872 Pseudomon
8	932.4	41.7	2904	1	SRFRPS	Y14317 Streptomyce
9	932	41.6	4910	1	AY040244	AY040244 Burkholde
10	912.4	40.8	10677	1	AE005967	AE005967 Caulobact
11	903.2	40.4	4754	1	RLE438039	AJ438039 Rhizobium
c	903	40.3	314100	1	SC0939106	AL939106 Streptomy
c	900.2	40.2	11677	1	AE011760	AE011760 Xanthomon
14	896.6	40.1	3027	1	AF126956	AF126956 Streptomy
15	890.8	39.8	298900	1	AP005937	AP005937 Bradyrhiz
16	881	39.4	3289	1	AF486647	AF486647 Rhizobium
c	879.6	39.3	295250	1	BX321862	BX321862 Nitrosomo
c	872.2	39.0	302150	1	AP001510	AP001510 Bacillus
c	871.6	38.9	11476	1	AE012219	AE012219 Xanthomon
20	865.6	38.7	2208	1	AB020234	AB020234 Bacillus
21	865.6	38.7	2232	1	AB020065	AB020065 Bacillus
22	865.6	38.7	2232	1	AB020120	AB020120 Bacillus
23	865.6	38.7	2235	1	AB020067	AB020067 Bacillus
24	865.6	38.7	2235	1	AB020092	AB020092 Bacillus
25	865.6	38.7	2238	1	AB020115	AB020115 Bacillus
26	865.6	38.7	2241	1	AB020094	AB020094 Bacillus
27	865.6	38.7	2247	1	AB020078	AB020078 Bacillus
28	865.6	38.7	2250	1	AB020106	AB020106 Bacillus
29	865.6	38.7	2253	1	AB020116	AB020116 Bacillus
30	865.6	38.7	2253	1	AB020117	AB020117 Bacillus
31	865.6	38.7	2256	1	AB020064	AB020064 Bacillus
32	865.6	38.7	2256	1	AB020066	AB020066 Bacillus
33	865.6	38.7	2256	1	AB020071	AB020071 Bacillus
34	865.6	38.7	2256	1	AB020072	AB020072 Bacillus
35	865.6	38.7	2256	1	AB020073	AB020073 Bacillus
36	865.6	38.7	2256	1	AB020074	AB020074 Bacillus
37	865.6	38.7	2256	1	AB020075	AB020075 Bacillus
38	865.6	38.7	2256	1	AB020076	AB020076 Bacillus
39	865.6	38.7	2256	1	AB020077	AB020077 Bacillus
40	865.6	38.7	2256	1	AB020079	AB020079 Bacillus
41	865.6	38.7	2256	1	AB020080	AB020080 Bacillus
42	865.6	38.7	2256	1	AB020081	AB020081 Bacillus
43	865.6	38.7	2256	1	AB020082	AB020082 Bacillus
44	865.6	38.7	2256	1	AB020083	AB020083 Bacillus
45	865.6	38.7	2256	1	AB020084	AB020084 Bacillus

ALIGNMENTS

RESULT 1
AR098264
LOCUS AR098264
DEFINITION Sequence 7 from patent US 6074860.
ACCESSION AR098264
VERSION AR098264.1 GI:12807521
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2238)
AUTHORS Robertson,D.E., Sanyal,I. and Adhikary,R.S.
TITLE Catalases
JOURNAL Patent: US 6074860-A 7 13-JUN-2000;
FEATURES Location/Qualifiers
LOCUS AR098264 2238 bp DNA linear PAT 14-FEB-2001

source	1. .2238	/organism="unknown"
BASE COUNT	634 a	545 c 605 g
ORIGIN	454 t	
Query Match	100.0%	Score 2238; DB 6; Length 2238;
Best Local Similarity	100.0%	Pred. No. 0;
Matches 2238; Conservative	0; Mismatches	0; Indels 0; Gaps 0
Qy	1	ATGAAATCAAAACACTCAGGATCTTCTACGTATAACAAACACTCGCGGAAATGC 60
Db	1	ATGAAATCAAAACACTCAGGATCTTCTACGTATAACAAACACTCGCGGAAATGC 60
Qy	61	CCTTTTACGGAGGTTCGCTTAAGCAAAGTCAGGTGGCGCACAAAACAGGATTGG 120
Db	61	CCTTTTACGGAGGTTCGCTTAAGCAAAGTCAGGTGGCGCACAAAACAGGATTGG 120
Qy	121	TGGCCCAACATGCTCAACCTCGGCATCTTACGCCAACATTATCGCTATCGACCCAAC 180
Db	121	TGGCCCAACATGCTCAACCTCGGCATCTTACGCCAACATTATCGCTATCGACCCAAC 180
Qy	181	GACCCGGATTTTGACTATGCCAAAGAGTTTAAAGACTAGATCTGGCAGCGTTAAAAAG 240
Db	181	GACCCGGATTTTGACTATGCCAAAGAGTTTAAAGACTAGATCTGGCAGCGTTAAAAAG 240
Qy	241	GACCTGCGAGCGCTTAATGACAGATTACAGACTTGGTGGCCAGCAGATTACGGTCATTAT 300
Db	241	GACCTGCGAGCGCTTAATGACAGATTACAGACTTGGTGGCCAGCAGATTACGGTCATTAT 300
Qy	301	GGCCCTCTTCTTATACGCATGGCGTGGCACAGCGCCGCACCTACCGTATCGGTGATGGC 360
Db	301	GGCCCTCTTCTTATACGCATGGCGTGGCNAGCGCCGCACCTACCGTATCGGTGATGGC 360
Qy	361	CGTGGTGGCGTGGCTCCGGCTCACAGGCTTCGCGCTCTCAATAGCTGCCACACAAT 420
Db	361	CGTGGTGGCGTGGCTCCGGCTCACAGGCTTCGCGCTCTCAATAGCTGGCCAGACAAT 420
Qy	421	GCCAACTGGATAAGCACGCTTCTCTTTTGGCCCATCAACAAAAATACGGTCGAAAA 480
Db	421	GCCAACTGGATAAAGCACGCTTCTCTTTTGGCCCATCAACAAAAATACGGTCGAAAA 480
Qy	481	ATCTCTCTGGGCGGACTCTAATGATACTCACAGGAAACGTAGCTCTGGAACCTATGGCTTT 540
Db	481	ATCTCTCTGGGCGGACTCTAATGATACTCACAGGAAACGTAGCTCTGGAACCTATGGCTTT 540
Qy	541	AAAACTTTTGGTTTTCAGGTGGCAGACAGATGTATGGGAGCCTGAAGAAGATGTATAC 600
Db	541	AAAACTTTTGGTTTTCAGGTGGCAGACAGATGTATGGGAGCCTGAAGAAGATGTATAC 600
Qy	601	TGGGAGCAGAAACCGAATGCTGGGAGACAAGCGCTATGAAGTGCACCGAGAGCTCGAA 660
Db	601	TGGGAGCAGAAACCGAATGCTGGGAGACAAGCGCTATGAAGTGCACCGAGAGCTCGAA 660
Qy	661	AATCCCTCTGGAGCCGTACAAATGGGACTCATCTATGTAAACCCGAGGACCCCAACGGC 720
Db	661	AATCCCTCTGGAGCCGTACAAATGGGACTCATCTATGTAAACCCGAGGACCCCAACGGC 720
Qy	721	AAGCCAGACCTTATCGCTGCTGCGCTGATATTTCGTGAGACTTTTGGCCGAATGGCAATG 780
Db	721	AAGCCAGACCTTATCGCTGCTGCGCTGATATTTCGTGAGACTTTTGGCCGAATGGCAATG 780
Qy	781	AATGACGAAGAACCGTGGCTCTCATAGCGGTGGACACACTTTCGAAAAACCCATGGT 840
Db	781	AATGACGAAGAACCGTGGCTCTCATAGCGGTGGACACACTTTCGAAAAACCCATGGT 840
Qy	841	GCTGCCGATCGGAGAAATATGTGGGCCGAGAGCCTGCCCGCAGAGTATTGAAGAAATG 900
Db	841	GCTGCCGATCGGAGAAATATGTGGGCCGAGAGCCTGCCCGCAGAGTATTGAAGAAATG 900
Qy	901	AGCCTGGGTGGAAAAACACCTACGGCACCGGACACGGTCCGGATACCATCACCAGTGA 960
Db	901	AGCCTGGGTGGAAAAACACCTACGGCACCGGACACGGTCCGGATACCATCACCAGTGA 960

Db 2041 GTTTTGAAGGAGAGACTTCAAAACTGCGAAGTAAGTGAAGTGGACCGGGTAGAC 2100
QY 2101 CTGATCTTCGGATCCAAATTCGAGCTTAAGAGCCCTCGCAGAAAGTGTACGGCTGTGCAGAT 2160
Db 2101 CTGATCTTCGGATCCAAATTCGAGCTTAAGAGCCCTCGCAGAAAGTGTACGGCTGTGCAGAT 2160
QY 2161 TCTGAAGAAAAGTTGTTAAAGATTTTGTGAAGGCTGGGCCAAAGTAATGAACCTGGAC 2220
Db 2161 TCTGAAGAAAAGTTGTTAAAGATTTTGTGAAGGCTGGGCCAAAGTAATGAACCTGGAC 2220
QY 2221 CGGTTTGATCTCAATAA 2238
Db 2221 CGGTTTGATCTCAATAA 2238

RESULT 2
AR215252
LOCUS AR215252 2238 bp mRNA linear PAT 25-SEP-2002
DEFINITION Sequence 7 from patent US 6410290.
ACCESSION AR215252
VERSION AR215252.1 GI:23313419
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 2238)
AUTHORS Robertson,D.E., Sanyal,I. and Adhikary,R.S.
TITLE Catalogues
JOURNAL Patent: US 6410290-A 7 25-JUN-2002;
FEATURES
source Location/Qualifiers
1..2238
/organism="unknown"
BASE COUNT 634 a 545 c 605 g 454 t
ORIGIN

Query Match 100.0%; Score 2238; DB 6; Length 2238;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGAAAATCACAAACACTCAGGATCTTCTAGCTATATACACAAACACTGGCGGAAATGC 60
Db 1 ATGGAAAATCACAAACACTCAGGATCTTCTAGCTATATACACAAACACTGGCGGAAATGC 60

QY 61 CTTTTACCGGAGTTTCGTTAAGCAAAAGTCAGGTGGCGGCACCAAAACAGGGATTGG 120
Db 61 CTTTTACCGGAGTTTCGTTAAGCAAAAGTCAGGTGGCGGCACCAAAACAGGGATTGG 120

QY 121 TGGCCCAACATGCTCAACCTCGGCATCTTACGCCAACATTCATCGCTATCGNACCCAAAC 180
Db 121 TGGCCCAACATGCTCAACCTCGGCATCTTACGCCAACATTCATCGCTATCGNACCCAAAC 180

QY 181 GACCCGGATTTTGACTATCGCGAAGAGTTTAAGAAGCTAGATCTGCGACGGTTAAAGAG 240
Db 181 GACCCGGATTTTGACTATCGCGAAGAGTTTAAGAAGCTAGATCTGCGACGGTTAAAGAG 240

QY 241 GACCTGGCAGCGCTAATGACAGATTCACAGGACTGGTGGCCAGCAGATTACGGTCAATTAT 300
Db 241 GACCTGGCAGCGCTAATGACAGATTCACAGGACTGGTGGCCAGCAGATTACGGTCAATTAT 300

QY 301 GGGCCCTCTTTATAGCATAGGGTGGGCACAGCGCGGCACCTACCGTATCGGTGATGGC 360
Db 301 GGGCCCTCTTTATAGCATAGGGTGGGCACAGCGCGGCACCTACCGTATCGGTGATGGC 360

QY 361 CGTGGTGGCGGTGGCTCCGGCTCACAGCGCTTCGGCCCTCTCAATAGCTGGCCAGACAAT 420
Db 361 CGTGGTGGCGGTGGCTCCGGCTCACAGCGCTTCGGCCCTCTCAATAGCTGGCCAGACAAT 420

QY 421 GCGAATCTGGATAAAGCAGCTTGCTTTTGGCCCATCAAAACAAAATACGGTCCGAAA 480
Db 421 GCGAATCTGGATAAAGCAGCTTGCTTTTGGCCCATCAAAACAAAATACGGTCCGAAA 480

QY 481 ATCTCTGGGCGGATCTAATGATFACTACAGGAACGTAGCTCTGGAAACTATGGGCTTT 540

Db 481 ATCTCTGGGCGGATCTAATGATFACTACAGGAACGTAGCTCTGGAAACTATGGGCTTT 540
QY 541 AAAACTTTTGTGTTTGCAGGTGGCAGCAGATGTATGGAGCCTCAAGAAAGATGTATAC 600
Db 541 AAAACTTTTGTGTTTGCAGGTGGCAGCAGATGTATGGAGCCTCAAGAAAGATGTATAC 600
QY 601 TGGGAGCAGAAACCGAAATGGCTGGGAGACAAAGCGCTATGAAGTGAACCGGACCGC 660
Db 601 TGGGAGCAGAAACCGAAATGGCTGGGAGACAAAGCGCTATGAAGTGAACCGGACCGC 660
QY 661 AATCCCTCGGAGCGGTACAAATGGACTCATCTATGTAACCCCAAGGACCCCAACGCG 720
Db 661 AATCCCTCGGAGCGGTACAAATGGACTCATCTATGTAACCCCAAGGACCCCAACGCG 720
QY 721 AAGCCAGACCTTATCGCTCGCTGATATTCGTGAGACTTTTGGCCGAAATGGCAATG 780
Db 721 AAGCCAGACCTTATCGCTCGCTGATATTCGTGAGACTTTTGGCCGAAATGGCAATG 780
QY 781 AATGACGAAGAAACCGTGGCTCTCATAGCGGGTGGACACACCTTCGGAACCAACATGGT 840
Db 781 AATGACGAAGAAACCGTGGCTCTCATAGCGGGTGGACACACCTTCGGAACCAACATGGT 840
QY 841 GCTGCCGATGGGAGAAATATGTGGSCCGAGAGCCTGCCGCCAGCTATTGAAGAATG 900
Db 841 GCTGCCGATGGGAGAAATATGTGGSCCGAGAGCCTGCCGCCAGCTATTGAAGAATG 900
QY 901 AGCTGGGTGGAAAACACCTACGCGCACCGGACGCTGGGATACCATCACCAGTGA 960
Db 901 AGCTGGGTGGAAAACACCTACGCGCACCGGACGCTGGGATACCATCACCAGTGA 960
QY 961 CTAGAAGCGCTGGACCAACCCCTACTCATGAGCAATAACTTTTTGAAAACCTC 1020
Db 961 CTAGAAGCGCTGGACCAACCCCTACTCATGAGCAATAACTTTTTGAAAACCTC 1020
QY 1021 TTTGTTAGAGTGGAGCTTACCAAAAGTCCAGCTGGAGCTTATCAGTGGAAACCAAA 1080
Db 1021 TTTGTTAGAGTGGAGCTTACCAAAAGTCCAGCTGGAGCTTATCAGTGGAAACCAAA 1080
QY 1081 GACGTTGCGGGGCTGGCACACCTACCGGATGCACATGATCCAGCAAGTGCACGCTCA 1140
Db 1081 GACGTTGCGGGGCTGGCACACCTACCGGATGCACATGATCCAGCAAGTGCACGCTCA 1140
QY 1141 TTTATGCTCACTACGACCTGGCGCTGGCATGGACCTGATTTACGAAAATTTCTCGA 1200
Db 1141 TTTATGCTCACTACGACCTGGCGCTGGCATGGACCTGATTTACGAAAATTTCTCGA 1200
QY 1201 CGGTACTATGAACCCCTGATGAGTTTGAGATGCTTTCGGAAGAGCTGTTACAACTG 1260
Db 1201 CGGTACTATGAACCCCTGATGAGTTTGAGATGCTTTCGGAAGAGCTGTTACAACTG 1260
QY 1261 ACACACAGATATGGGACCAAAAGTTCGCTAGCTGGGACCAAGTGCCTCAGGAAGAC 1320
Db 1261 ACACACAGATATGGGACCAAAAGTTCGCTAGCTGGGACCAAGTGCCTCAGGAAGAC 1320
QY 1321 CTCATCTGGCAGACCTTATACAGATGTAAGCCATCCTCTTGTAGACGAAAACGATATT 1380
Db 1321 CTCATCTGGCAGACCTTATACAGATGTAAGCCATCCTCTTGTAGACGAAAACGATATT 1380
QY 1381 GAAGGCTTAAAGCCAAAATCCTGGAAATCGGAGCTGACGGTAAAGCGAGCTGGTAAGCAG 1440
Db 1381 GAAGGCTTAAAGCCAAAATCCTGGAAATCGGAGCTGACGGTAAAGCGAGCTGGTAAGCAG 1440
QY 1441 GCATGGGCTCTCATCTACTTTTAGAAACTCTGACAAGCGCGCGGTGCCACCGTGCA 1500
Db 1441 GCATGGGCTCTCATCTACTTTTAGAAACTCTGACAAGCGCGCGGTGCCACCGTGCA 1500
QY 1501 CGTATACGACTGGCCCCACAAAAGACTTGGAAAGTAAACAACTCAGCAACTTGCAGG 1560
Db 1501 CGTATACGACTGGCCCCACAAAAGACTTGGAAAGTAAACAACTCAGCAACTTGCAGG 1560
QY 1561 GTACTCAAAACACTAGAAGGTATCCAGGAGGACTTTAAACAGCGGCATCAGATAACAA 1620
Db 1561 GTACTCAAAACACTAGAAGGTATCCAGGAGGACTTTAAACAGCGGCATCAGATAACAA

Db	1561	GTACTCAAAACACTAGAGGTTATCCAGGAGGACTTTAACAGGCGCAATCAGATAACAAA	1620
QY	1621	GCAGTATCGTTGGCCGACCTGATTGTGCTGCGCGCTGTGCGGCTGTAGAAAAGCTGCA	1680
Db	1621	GCAGTATCGTTGGCCGACCTGATTGTGCTGCGCGCTGTGCGGCTGTAGAAAAGCTGCA	1680
QY	1681	AAAGATCGTGGCCATGAGTGCAGGTGCGCTTTTCAACCGGACGAGCGGATGCCACCGCT	1740
Db	1681	AAAGATCGTGGCCATGAGTGCAGGTGCGCTTTTCAACCGGACGAGCGGATGCCACCGCT	1740
QY	1741	GAGCAAAACCGATGTGGAAGCTTTTGAAGCACTAGAGCGAGCGGCTGACGGCTTTAGAAAC	1800
Db	1741	GAGCAAAACCGATGTGGAAGCTTTTGAAGCACTAGAGCGAGCGGCTGACGGCTTTAGAAAC	1800
QY	1801	TACATTAAACCGGAGCAATAAGTATCGCTGAGGAAATGCTCGTAGACCGGCGCAGCTT	1860
Db	1801	TACATTAAACCGGAGCAATAAGTATCGCTGAGGAAATGCTCGTAGACCGGCGCAGCTT	1860
QY	1861	CTGTCGCTTTCCGACCAAGAAATGCTGCTTTGCTAGCGGTATGCGTGAATGCGGCACC	1920
Db	1861	CTGTCGCTTTCCGACCAAGAAATGCTGCTTTGCTAGCGGTATGCGTGAATGCGGCACC	1920
QY	1921	AACATACGACGGTTCGACGACGAGTGTGTTTACAAATAGCGCGGTGACGATATCCAATGAC	1980
Db	1921	AACATACGACGGTTCGACGACGAGTGTGTTTACAAATAGCGCGGTGACGATATCCAATGAC	1980
QY	1981	TTCTTTGTAACCTGCTAGACCTCAACACTAAATGGCGAGCGCAGATGAATCAGACAAA	2040
Db	1981	TTCTTTGTAACCTGCTAGACCTCAACACTAAATGGCGAGCGCAGATGAATCAGACAAA	2040
QY	2041	GTCTTTGAAGCGCAGACCTTCAAACTGGCGAAGTAAAGTGGAGTGGCACCCTGGGTAGAC	2100
Db	2041	GTCTTTGAAGCGCAGACCTTCAAACTGGCGAAGTAAAGTGGAGTGGCACCCTGGGTAGAC	2100
QY	2101	CTGATCTTCGGATCCAAATTCGAGCTTAAGAGCCCTCGCAGAAAGTGTACGGCTGTGCAGAT	2160
Db	2101	CTGATCTTCGGATCCAAATTCGAGCTTAAGAGCCCTCGCAGAAAGTGTACGGCTGTGCAGAT	2160
QY	2161	TCTGAAGAAAGTTGTTAAAGATTTGTGAAGCCCTGGGCGCAAGTAATGGACCTGGAC	2220
Db	2161	TCTGAAGAAAGTTGTTAAAGATTTGTGAAGCCCTGGGCGCAAGTAATGGACCTGGAC	2220
QY	2221	CGGTTTCATCTGAATAA 2238	
Db	2221	CGGTTTCATCTGAATAA 2238	
RESULT 3			
LOCUS	AE010762	11320 bp	DNA linear BCT 03-APR-2002
DEFINITION	Methanosarcina acetivorans str. C2A, section 107 of 534 of the complete genome.		
ACCESSION	AE010762 AE010299		
VERSION	AE010762.1 GI:19914790		
KEYWORDS			
SOURCE	Methanosarcina acetivorans C2A		
ORGANISM	Methanosarcina acetivorans C2A		
REFERENCE	Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis; Methanosarcinales; Methanosarcinaceae; Methanosarcina.		
AUTHORS	Galagan, J.E., Nusbaum, C., Roy, A., Endrizzi, M.G., Macdonald, P., FitzRugh, W., Calvo, S., Engels, R., Smirnov, S., Atnoor, D., Brown, A., Allen, N., Naylor, J., Stange-Thomann, N., DeArellano, K., Johnson, R., Linton, L., McEvan, P., McKernan, K., Talamas, J., Tirrell, A., Ye, W., Zimmer, A., Barber, R.D., Cann, I., Graham, D.E., Grahame, D.A., Guss, A., Hedderich, R., Ingram-Smith, C., Kuettner, C.H., Krzycki, J.A., Leigh, J.A., Li, W., Liu, J., Mukhopadhyay, B., Reeve, J.N., Smith, K., Springer, T.A., Umayam, L.A., White, O., White, R.H., de Macario, E.C., Ferry, J.G., Jarrell, K.F., Jing, H., Macario, A.J.L., Paulsen, I., Pritchett, M., Sowers, K.R., Swanson, R.V., Zinder, S.H., Lander, E., Metcalf, W.W. and Birren, B.		
TITLE	The Genome of M. acetivorans Reveals Extensive Metabolic and Physiological Diversity		
JOURNAL	Genome Res. 12 (4), 532-542 (2002)		
MEDLINE	21929760		
PUBMED	11932238		
REFERENCE	2 (bases 1 to 11320)		
AUTHORS	Birren, B.		
TITLE	Direct Submission		
JOURNAL	Submitted (20-MAR-2002) Center for Genome Research, Whitehead Institute, Nine Cambridge Center, Cambridge, MA 02141, USA		
FEATURES	Location/Qualifiers		
source	1..11320		
	/organism="Methanosarcina acetivorans C2A"		
	/mol_type="genomic DNA"		
	/strain="C2A"		
gene	/db_xref="taxon:188937"		
	303..695		
CDS	/gene="MA0965"		
	303..695		
	/gene="MA0965"		
	/codon_start=1		
	/transl_table=11		
	/product="thioredoxin"		
	/protein_id="AM04398.1"		
	/db_xref="GI:19914791"		
	/translation="MDENYVEIEDATWGQOVEDSEKPVIVMFYSPCTPCYKAEHPV		
	AEYAEKRAFAIRINIIWNPMTAERYGVQGTPTFKFCHGKQWQVQGIYPSILG		
	NAIRDMIQYGEICIRKTPVGDITGV"		
gene	complement(934..1380)		
CDS	/gene="MA0966"		
	complement(934..1380)		
	/gene="MA0966"		
	/codon_start=1		
	/transl_table=11		
	/product="predicted protein"		
	/protein_id="AM04399.1"		
	/db_xref="GI:19914792"		
	/translation="MTYFELNGLALNWLGMVTVAYSTIRNVELKMKLLLPVI		
	LSFTIANLVLYNFEVFDLGNLEARHLPIGSLNLSRGNIVGKNLTRISAEM		
	KIDFTASLAGASRYEALMPYTGNAFTPHSLRSLPILQQLSLSPG"		
gene	complement(1732..2238)		
CDS	/gene="MA0967"		
	pseudo		
	2676..3629		
gene	/gene="MA0968"		
CDS	2676..3629		
	/gene="MA0968"		
	/codon_start=1		
	/transl_table=11		
	/product="eif5A-deoxyhypusine synthase"		
	/protein_id="AM04400.1"		
	/db_xref="GI:19914793"		
	/translation="MDPEASARDLTPVKAKIVPNMSVDLREYAGCAFGAGRLAE		
	AVDIYEMLASERTTKFFGLAGAMTPAGMRGIIADLIRGDHIDVLTGANNVHDTVE		
	ALGLHHYKSDCANDIOLRHECIDRIYDYLPOHFTDLEFLQSVTAGLPQEKLSIR		
	QVLETKNLDSSSILKTAEMGVPCPALQDSVIGLQAWLYKEGNLRLHVDAPADM		
	HEFMEICYAEASAGAMLLGGVPKNYILQMLVTPKSFDAIQLTMDRPTGGTGGAT		
	LDEAQSMGKVGEDAKSVTVYADATITLPLIVSAVRTRLSKR"		
gene	3568..4299		
CDS	/gene="MA0969"		
	3568..4299		
	/gene="MA0969"		
	/codon_start=1		
	/transl_table=11		
	/product="orotidine-5-phosphate decarboxylase"		
	/protein_id="AM04401.1"		
	/db_xref="GI:19914794"		
	/translation="MQQSLSHLSPFLCGRAFQRDDLMERNCTMLALDVTDRREALI		
	AEDWVEFVDAIKVGYPLIATGIIIRLAEFAPVIADFKVADIPNTNRLICQVFEA		
	GADAVTAQGGTGRSDIACIEASVKDFVVSMSHPGGADFMQVAGAIKMALE		
	AGAGLVAPATPRKVRKIRRIIGDKLTIIISPGVGAQGGKASDVIAAGADWIVGRSI		
	YKASPREAARQIAEEIQALRGEY"		
gene	5307..7799		
CDS	/gene="MA0970"		
	5307..7799		


```

/gene="MA0970"
/codon_start=1
/transl_table=11
/product="sensory transduction histidine kinase"
/protein_id="AA04402.1"
/db_xref="GI:19914795"
/translation="MQEYSKINSWQKQKGNMNAKVQGTATNQNVLVSDNNGTI
INSNASEPILKMGVSGEKLFPSSLDVLRVISRNSPEKTEVKGNDLYLVFVSP
PEQVRVSDISDQKEFKMGVSETREKAEDTLVEELRSEKRYRTLFTDIDSGF
CIEVFLDADCPIDYRLEINAFEROTGKNAVGRMRREIAPEHEHWFQIYGNVA
LTGEPVRFNPAALFGLHFDYAFRTGDPDHRVGLFNDLAKRKAEEMLRESAEO
AFILQSLDALRPLDPIEQAVATRVLGEHLGARVAYGEVTDDDTFIVFERNVAPG
VLVLTGFRFMADEGSSLNALKEGRTIVVPDLASSTELSEARSGYANGLIASLMGP
LVKGRFVAGLVVHHTPRQMSAEVALLIEETAEARTWAAVERAKAEALKESEERT
LSTSPISGVSAASEKIIITNPAYDRILGNPSELIGRNSDTPVYNPEERKSWCRTL
KQGVYRDETFFKKKDGTPVWYSINTSPISFGGRKAAMGTIODISERKQAEALRES
ELKFRFLSTLLEKKEVTELEKAYFLQEIDIIRKQETHHRIKNNLOVISLDDLO
AEKFRCKNIEDSKILEAFKESODRVISMLIHEELHKSGETDPLNFSAYIHELNSNL
FLSYRLNGDISLMDIEDIDPDMDTSPVPLGMVNELVNSLKHAFSDRDKGEIRIK
LHRKKNRESIEDCCAFILSDNGIGIPKLEIEDIESLGLQLVTLTLLDQJGLE
LKKDGTETIVRFSITERDNOANSPTQQPNEF"
/gene="fur"
/notes="synonym: MA0971"
/complement(8223..8672)
/gene="fur"
/codon_start=1
/transl_table=11
/product="ferric uptake regulation protein"
/protein_id="AA04403.1"
/db_xref="GI:19914796"
/translation="MSNIQDSPLIILKALRGKYKATPORIATGQVFLNNHHPAQR
IYSEKVKVPTVSLATYIKTVQILKEVLIQEFNLKQDARFDPNPEHAHLVCLKQK
SINDCTDPMISEIVLRMSNEVDFSGAEWNFINFGICNCRKKSIQD"
8913..11123
/gene="MA0972"
8913..11123
/gene="MA0972"
/codon_start=1
/transl_table=11
/product="catalase/peroxidase"
/protein_id="AA04404.1"
/db_xref="GI:19914797"
/translation="MGNVNTDDKMSVTSKANQETGRDMSNRDWPNNILKLEILHQ
HSSKPMGDEDFNAYEFKSLDLAAVKKDLAALMDSQDWWPADFGHYGLFIRMAWH
SAGTYRAGRGGGGGRQQRFLNPNNDKARLLPDKRYGRKISWADLMI
LTGNVAMETMFTFGGGREDVDPDQDVWGEDTDLGDERYTGDRDLENPLAAY
OMGLIYVPEGPNGNDPIAAAKDIREVFARMAMDBETVALLIAGHAFCKTHGAPV
SHVGPEEASIEAOGLWKSSPCTCKGDDTITGGLEVTVNTPTTWSNNFFRILGY
EWELTSPAGAYQWPKGGAGCTIPDAHPDSRHAPSMMITDLSLRFDPVYBKISHH
FYENPQOLADSFARAFKLTNRDMGPRARYLGPVPAEELIWQDDPIPAVNHILIDED
IAFLKDRILASGLISIQLYSTAWVSASTFRGSDRGANGARILAPOKDWEYNQPAE
LAKVLNTLEGIOSEFNSAASGGKKVSLADILVLGAGGVQBAAKNAGYDVTVPFLGR
MDALQEDTVVSPFALLEPDAQFNLYKAOYPLPAEBELLVDKAQLLTLTAPEMTVLGV
GMRLVNTFGHTOHGVFTOKPEALTNDFFVNLDMGTENKAVSDVKDIPGCDKRTGE
VKMTGRVDLIFGNSQLRALAEVYGSADAQEFVQDFVAAATKVMNLDLFDLA"
BASE COUNT 3077 a 2495 c 2699 g 3049 t
ORIGIN
Query Match 47.1%; Score 1053.2; DB 1; Length 11320;
Best Local Similarity 68.2%; Pred. No. 2.6e-273;
Matches 1479; Conservative 0; Mismatches 688; Indels 3; Gaps 1;
Qy 62 CTTTATACGGAGGTTGCGTTAAGCAAGTGCAGGTGGCGGCACCAAAACAGGAGTTGGT 121
Db 8950 CTGTAACGACGCGAGCAACAAACAGAACCGGTAGAGACATGTCGAACCGCGACTGGT 9009
Qy 122 GGCCCAACATGCTCAACCTCGGATCTTACGCCAACATTCATCGCTATCGGACCCCAACG 181
Db 9010 GGCCGACCACTTGAAGCTTTGAGATCTCGCACCAGCATTCCTCCAAGTCCAACCGATGG 9069
Qy 182 ACCCGGATTTGACTATGCCGAGAGCTTTAAGAGCTAGATCTGGCAGCGGTTAAAAAG 241

```

```

Db 9070 GCAGGAGCTTCAATACGCGAAGAGGTTCAAGAGCCCTAGATCTGGCGGCTGTGAAAAAG 9129
Qy 242 ACCTGGCAGCGCTAAATGACAGATTTCAGGACTGGTGGCCAGCAGATTAACGGTCAATTATG 301
Db 9130 ATCTTGGCGCGCTGATGACCGGATTCGAGACTGGTGGCCGGGGAATTTTGGCCATTATG 9189
Qy 302 GCCCTTCTTTATACCGATGGCGTGGCAGAGCGCGGACCTTACCCTATCGGTATCGGTATG 361
Db 9190 GCCCTTGTGTTATTCGATGGCATGGCAGCAGCGAGCCTTATCGCGCGGTGACGCGCC 9249
Qy 362 GTGGTGGCGGTGGCTCGCGCTCACAGGCTTCGCGCCTCTCAATAGCTAGCTGCCAGACAAT 421
Db 9250 GCGCGCGCGCGGTAGAGGCCAGAACGCTTCGCGCGCGCTCAACAGTTGGCCGCAATG 9309
Qy 422 CCAATCTGGATAAAGCAGCGCTTGCTTTTGGGCCCATCAACAAAAATACGGTGCAGAAA 481
Db 9310 TCAACCTGGACAAAGCGCGCTGCTGCTCTGGCCGATCAAGCAGAGAATGATGGCGGAAA 9369
Qy 482 TCTCTGGGCGGATCTAATGATACTCACAGAAACGTAGCTCTGGAACCTATGGGCTTTA 541
Db 9370 TTTTCATGGGCGGACCTTATGATCTTACGGGCAACGTCGCAATGGAACTATGGGATTC 9429
Qy 542 AAATTTTGTGTTTTCAGGTGGCAGAGCAGATGTATGGAGCCCTGAAGAAGATGTATACT 601
Db 9430 AAACGTTTGTGTTTTCGGGGCGGTCGCGAGGACGCTCTGGAGCCGCGACGACCTCTATT 9489
Qy 602 GGGGAGCAGAAACCGGATGGCTGGGAGACAAGCGCTATGAAGGTGACCGGAGAGTCGAAA 661
Db 9490 GGGGCTCTGAGGACACGCTGGCTGGTGATGACGCTACACGGGCGCGGAGACCTCGAAA 9549
Qy 662 ATCCCTGGGAGCGGTACAAATGGGACTCATCTATGTATAACCCCGAAGGACCCAAACGGCA 721
Db 9550 ATCCGCTCGCTGCGGTGCGATGAGGTGGGCTGATCTATGTCAATCCGGAAGGCGCAACGCA 9609
Qy 722 AGCCAGACCTCATCGCTGCTGCGGTGATATCTGTGAGACTTTTGGCCGGAATGCAATGA 781
Db 9610 ATCCGATCGATAGCGCGGCGCAAAATATCCGCGAGGTTTTCGCTCGCATGCCCATA 9659
Qy 782 ATGACGAAGAACCGGTGGCTCTCATAGCGGTGGACACACCTTCGGAACCCATCGATGGT 841
Db 9670 ACGAGGAAGAGACGTTTGGCTCATCTGCTGCGGTGTCACGCTTCGGAACCCACCGTG 9729
Qy 842 CTGCGGATCGGAGAAATATGTGGCGGAGAGCCTGCGCGCGAGGTATTTGAAGAAATGA 901
Db 9730 CCGGCGCTGC---GTCCCATGTGCGGCTGAGCCAGAAAGCGGCAGTATCGAGCGCAAG 9786
Qy 902 GCCTGGGTTGAAAAACACCTACGGCACCGGACGCTGGGATACCATCACCACTGGAC 961
Db 9787 GTCTCGTTGGAAGAGCAGCTTCGGCACTGGCAAGCGGATGACAGCATACCGCGCGCC 9846
Qy 962 TAGAAGCGCTGGACCAAGACCCCTACTCAATGGAGCAATACTTTTTTGAACACCTCT 1021
Db 9847 TGGAGGTCACTTGGACCAACACACCAACGAAGTGGAGTAACAACCTTCTTCAGGATCTGT 9906
Qy 1022 TTGGTTACGAGTGGGACCTTACCAAAAGTCCAGCTGGAGCTTATCATGTGGAACCAAAAG 1081
Db 9907 TCGGCTACGAGTGGGAATTGACGAAGAGCGCGCGGTGCTATATCATGTGGAAGCTTAAG 9966
Qy 1082 ACGGTGCGGCGGTGGCACCATACCGGATGCACATGATCCAGCAAGTCCGACCGCTCCAT 1141
Db 9967 GTGCGGAGGTGCGGCGACCATTCGCGACGCCACAGACCTTCCAGCGTCCAGCGGCAA 10206
Qy 1142 TTATGCTCACTACGGACCTGCGCTGCGCATGGACCTGTATGAGAAAAATTTTCTGCAC 1201
Db 10027 GCATGATGACCACTGACCTTTCTTTCCTTTCGAGTTTCGACCCCGCTACGAAAAAGATTTCAAGGC 10086
Qy 1202 GGTACTATGAACCCCTGATGAGTTTGCAGATGCTTTTCGCGAAAGCATGGTACAACTGA 1261
Db 10087 ACTTCTACGAGAACCCGGGTCAATTAGCGGACTCATTCGCGCGGCGCATGGTTCAAGTGA 10146
Qy 1262 CACACAGATATGGGACCAAAAGTGGCTTACCTGGGACCAAGAGTGCCTCAGGAAGACC 1321
Db 10147 CTCACCGGACATGGGTCCGCGCGCGCTATCTTCGCTCGGAGGTTCTCTGCCGAAGAGC 10206

```

KEYWORDS		Ralstonia solanacearum
SOURCE		Ralstonia solanacearum
ORGANISM		Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae; Ralstonia.
REFERENCE		1
AUTHORS		Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S., Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L., Chandler,M., Cholsne,N., Claudel-Renard,C., Cunnac,S., Demange,N., Gaspin,C., Layle,M., Moisan,A., Robert,C., Saurin,W., Schlex,T., Sigulier,P., Thebaud,P., Whalen,M., Wincker,P., Levy,M., Weissenbach,J. and Boucher,C.A.
TITLE		Genome sequence of the plant pathogen Ralstonia solanacearum
JOURNAL		Nature 415 (6871), 497-502 (2002)
MEDLINE		21681879
PUBMED		11823852
REFERENCE		2 (bases 1 to 198050)
AUTHORS		Bucher,C.A.
TITLE		Direct Submission
JOURNAL		Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
COMMENT		Cremlieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie Moleculaire des Interactions Plantes-Microorganismes INRA-CNRS, BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS 118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA URGV, 2 rue Gaston Cremlieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27, F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire INRA, BP27, F31326 Castanet-Tolosan Cedex
FEATURES		Christian.Bucher@toulouse.inra.fr
source		http://sequence.toulouse.inra.fr/R.solanacearum.html.
FEATURES		Location/Qualifiers
gene		1..198050
CDS		/organism="Ralstonia solanacearum"
		/mol_type="genomic DNA"
		/strain="GMI1000"
		/db_xref="taxon:305"
		151..642
		/gene="Rsc0770"
		/note="synonym: RS05084"
		151..642
		/gene="Rsc0770"
		/EC_number="1.-.-.-"
		/function="small molecule metabolism"
		/note="Product confidence : putative
		Gene name confidence : hypothetical
		predicted by Codon_usage
		predicted by Framed"
		/codon_start=1
		/evidence=not_experimental
		/transl_table=11
		/product="PUTATIVE TRANSMEMBRANE DEHYDROGENASE (SMALL SUBUNIT) OXIDOREDUCTASE PROTEIN"
		/protein_id="CAD14472.1"
		/db_xref="GI:17427782"
		/db_xref="GOA:Q8YIC0"
		/db_xref="SPTREMBL:Q8YIC0"
		/translation="MHTPPSGDDPAGLIRROWLQSLAATAASLVGSVTLVAVAO
		APGEPDLAFMGLSQALTARSALDRVAGRLLAALQOASAGFAAOLRPLAQSIACTLS
		GLQOETALRLIQAWYTGVDGTVTYEQALMGVVDILVITFCPNRPGFWAEPPEV
		KQS"
		642..2261
		/gene="Rsc0771"
		/note="synonym: RS05083"
		642..2261
		/gene="Rsc0771"
		/EC_number="1.-.-.-"
		/function="small molecule metabolism"
		/note="Product confidence : putative
		Gene name confidence : hypothetical
		predicted by Codon_usage
		predicted by Homology
		predicted by Framed"
gene		
CDS		
1322	TCATCTGGCAGACCCCTATACAGATGTAAGCCATCTCTTGTAGACGAAACGATATTTG	1381
10207	TCATCTGGCAGACCCCAATCCCGGGTCAATCAGAGTTGATCGATGAGAGGATATTTG	10266
1382	AGGCGCTAAAGCCAAATCTTGGATCGGGAATCAGGTAAGCGAGCTGTGAACACGG	1441
10267	CTTCTCAAGCAGAGGATCTTGCTCAGGCTGTCGATCTCGCAACTGGTTTCGACGG	10326
1442	CATGGCTTCTCATCTACTTTTGAACACTGGAAGCGCGGGTGCACACGGTGGAC	1501
10327	CCTGGGTGCGGCTCCACCTTCCGTGCTCGCAACAGTGGCGGTGCAACCGCGGAC	10386
1502	GTATACGACTGGCCCCCACAAGAACTGGGAAGTAACAACCCCTCAGCAACTTGCAGG	1561
10387	GAATTCGGCTCGCGCGCAGAGGATGGGAAGTCAACACCGCCGAAGTGGCGAAG	10446
1562	TACTCAAAACACTAGAAAGTATCAGGAGGACTTTAACCGCGCAATCAGATAACAAG	1621
10447	TGCTGAACACTCTCAGGGTATCCAGAGCGAGTTCAACAGCGCGCTCCGCTGGCAAG	10506
1622	CAGTATCTGTGGCCGACCTGATGTGCTGGCGGCTGTGCGGCTGTAGAAAAGCTGCAA	1681
10507	AGGTATCGCTTCTCACCTGATGTTCTGGCTGTTGTGCGGCTGCGAGCGCTGCGA	10566
1682	AGATGCTGGCCATGAGTGCAGGTGCTTCAACCCGGGAGCGGCGATGCCACCGTGG	1741
10567	AGATGCTGGTACGATGATGATGTTCTTCTTGGCGGCGCATGATGCTTGCAGG	10626
1742	AGCAAAACGATGTGAAGCTTTTCAAGCACTAGAGCCAGCGGCTCACGCGCTTTAGAACT	1801
10627	AGCAGCCGATGTTCT	10686
1802	ACATTAACCCGAGCATAAAGTATCCGCTGAGGAAATGCTGTAGACGGCGGCGAGCTTC	1861
10687	ATCTCAAGGCCAGTACCCCTTACCGCGCGAGAGTTCGTGTTTCAACAGCGCAATTGC	10746
1862	TGTCCTTTGGCACCACCAATGACTGCTTGGTATGGCGGTATGCGGTACTGGGCACCA	1921
10747	TGACTCTGACCGCACGAGATGAGGCTTGTGTTGGCGCATGCGCGTGTGTAACACCA	10806
1922	ACTACGCGGTTGCGCATGATGAGTGTTCACAAATGAAGCGGTCAGCTATCCATGACT	1981
10807	ACTCGGACATACCCAGCAGCGCTTTTACCAGACCGGAAGCGCTTAACCAACGACT	10866
1982	TCTTTGTAACCTGTAGACCTCAACACTAATGGCGAGCCAGCGATGAATCAGACAAAG	2041
10867	TCTTCTGAACTCTGACATGGTACGGAATGAAGCGCTATCGGACGTCAGGACA	10926
2042	TTTTTGAAGGACAGACTTCAAACTGGGGAAGTAAAGTGGAGTGGCACCAGCGGTAGAC	2101
10927	TATTTGAAGGTCGATCGCAAGACTGGCGAAGTCAAGTGAAGTGGACCGGCGCGTGCATC	10986
2102	TGATCTTGGATCCAAATTCGAGCTTAAGACCCCTCGCAGAGCTGTACCGCTGTGCAGAT	2161
10987	TCATCTCGGTTCTAATCTCCAGCTCCGGCTCTTCCGAGGTCTACCGAAGCGCGATG	11046
2162	CTGAAGAAAGTTTGTAAAGATTTTGAAGGCTCGGCCCAAGTAATGACCTGGAC	2221
11047	CGCAGGAAGTTTCTTCCAGGACTTCTGTCAGGCTTCTGTCGCGGCTCGGACCAAGTA	11106
2222	GTTTGTATCT	2231
11107	GCTTCGACCT	11116
RESULT 4		
AL646061/c		198050 bp DNA linear BCT 24-JAN-2003
LOCUS		Ralstonia solanacearum GMI1000 chromosome, complete sequence;
DEFINITION		segment 5/19.
ACCESSION		AL646061
VERSION		AL646061.1 GI:17427781

Qy	390	CTTCGCGCCTCTCAATAGCTGGCCAGACAATGCCAATCTCGATAAGCAGCGCTTGCTTCT	449
Db	8725	CTTCGCGCGCTCAACAGCTGGCCGACAAAGCGCAACCTCGACAAGCGCGCGCTGCT	8666
Qy	450	TTGCGCCATCAAAACAAAATACGGTCGAAAAATCTCTCTGGCGGATCTAATGATACTCAC	509
Db	8665	GTGCGCCATCAAGCAGAAGTACGGCCGCAACATTTCTGTGGCGCGACCTCATGATCTCTCAC	8606
Qy	510	AGAAACGTAGCTCTGGAACATATGGCTTTAAACATTTTGGTTTTCAGAGTGGCAGAGC	569
Db	8605	GGGCAAGCTCGCGTCGAGTCGATGGGCTTCAAGACCTTCGGCTTCGCCGCGGCGCAA	8546
Qy	570	AGATGATGGAGCCTCAAGAGAATGTATCTCTGGGAGCAAAACCAATGGCTGGGAGA	629
Db	8545	GGATGATGGGAACCGGAAGAGGATGTCTACTGGGTTCCGAAACCACTTGGTGGGCGA	8486
Qy	630	CAAGCGCTATGAAGGTACCGAGAGCTCGAAAATCCCTCTGGGAGCCGTACAAATGGGACT	689
Db	8485	CCAGCGCTACACCGCGCATCGGATCTGGAGACCCGCTCCGCGGTCCAGATGGCGCT	8426
Qy	690	CATCTATGTAAACCCCGAAGAGCCCAACGGGAGCCAGACCTTATCGCTGCTGGCGGTGA	749
Db	8425	GATCTAGTCAATCCGGAAGGCCGGAACGGCAACCCCGATCCAATCGCGCGCCAGGGA	8366
Qy	750	TATTCGTGAGACTTTTGGCGGAATGGCAATGAATGACGAAGAAACCTTGCTCTCATAGC	809
Db	8365	CATCCGGGAGACTTTCGCCCGCATGGCGATGAACGATGAAGAGACGGTTCGCGTGAATCGC	8306
Qy	810	GGGTGGACACACTTCGGAAAAACCCATGTGTGTCGATGCGGAGAGAAATATGTGGCGCG	869
Db	8305	GGGCGGCACCTCTTCGGCAAGACCCACGGGCGCG----GTCCGGCTCCAAGCTGGGCGC	8250
Qy	870	AGACCTTCGCGCCGACGGTATGTGAAGAAATGAGCCTGGGTGGGATGGAAACACCTACGCGAC	929
Db	8249	GGAGCGGAAGCGCGCGGCATCGAAGAAACAGGGCCTGGGCTGGAGCAGCGCTTCGGCAC	8190
Qy	930	CGGACACGGTCGGATACCATCACCACTGAGCTAGAGGCGCTGGGACCAAGACCCCTAC	989
Db	8189	CGGCAAGGCACCGCACGCCATCACACGGCGCTGGAAAGTGAACCTGGACCAACCGCCGAC	8130
Qy	990	TCAATGAGGCAATAACTTTTTTGAACACCTCTTTGGTTACGAGTGGGAGCTTACCAAAG	1049
Db	8129	GCAGTGAGCAACAACCTCTTCCAGAACCTGTTTCGGCTACGAGTGGGAACTGACCAAGAG	8070
Qy	1050	TCCAGCTGGAGCTTATCAGTGGAAACAAAGACGGTCCGGGCTGGCACCATACCGGA	1109
Db	8069	CCGGCGCGCGCACACAGTGGGTGCGCAAGGACGCGCGCGAG-----ACCATTCCGGA	8016
Qy	1110	TGCACATGATCCAGCAAGTCGACGCTCCATTTATGCTCACTAGGACCTGGCGTGGC	1169
Db	8015	TGCGCACGACCGTCCCGGAAGCATCTGCGCAGATGCTGACCAACCGACCTGTCCCTGGC	7956
Qy	1170	CATGGACCCCTGATTACGAAAAATTTCTCGACGGTACTATGAAACCCCTGATGAGTTGC	1229
Db	7955	CTTCGACCCCGCTACGAGAAGATCTCACGGCGCTTCTACGAGCATCCGAGCAGTTCCG	7896
Qy	1230	AGATGCTTTCCGGAACCATGTTACAACCTGACACAGAGATATGGGACCAAGGTGGC	1289
Db	7895	CGAGCGCTTCGCGCGGCGTGGTTTCAAGCTTGACCCACCGGACATGGGCGCGCGCGC	7836
Qy	1290	CTACTGGGACCAAGTGCCTCAGGAAGACCTCATCTGGCAAGACCCCTATACCATGAT	1349
Db	7835	CTATCTCGGCGCGGAAGTACCGGCGGAAGCGCTCATCTGCGAAGACCCGATCCCGCCGT	7776
Qy	1350	AAGCCATCCTCTTGTAGACAAAACGATATGTGAAGGCCTAAAAACCCAAATCTTGGAAATC	1409
Db	7775	GGATCAACAGCTGATCGACACGCGAGGACATTCGCGCCCTCAAGGCCAGGATCTGGCTTC	7716
Qy	1410	GGGACTCACGGTAAGCAGCTGGTAACACCGCATGGGCTTCTGATCTACTTTTAGAAA	1469
Db	7715	GGGGCTTCGGCTTCGCAACTGGTCTCGACCGCTGGGATCGGGCTCACTTCCGCGG	7656

QY	1470	CTCTGACAAAGCGCGCGGTGCCACACGCTGACAGTATACGACTTGCCGCCCAAAAAGACTG	1529
Db	7655	CTCGGACATGCGCGCGCGGTGCCAAACGCGCGCGCATCCGTCTCGCGCGCGAGAGGATTTG	7596
QY	1530	GGAAGTAAACAACCCCTCAGCAACTTGCCAGGTACTCAAACACTACGAAGTATCCAGGA	1589
Db	7595	GGAAGCCAAACGCGCGCGCATGCGTGTGCGCAAGAAGTCTCGCTCGCCGACCTGATCGTGCT	7536
QY	1590	GGACTTTAAACGAGCGCAATCAGATAAACAAGCAGTATCGTTGGCCGACCTGATTTGCT	1649
Db	7535	CGCGTTCAACGCGCGCAGTCTGCTGTGCGCAAGAAGTCTCGCTCGCCGACCTGATCGTGCT	7476
QY	1650	GGCCGCGTGTGCGGGGTAGAAAAGCTGCMAAAGATGCTGGCCATGAGGTGCAGGTGCC	1709
Db	7475	GGCCGCGTGTGCGCGGCTCGAAACAGGAGCGGAAGAACCGGCGCATCGGTGGAGGTGCC	7416
QY	1710	TTTCAACCCGGGAGAGCGGATGCCACCGCTGAGCAAAACCGATGTGGAAGCTTTCGAAGC	1769
Db	7415	CTTCACGCGGGGCGCATGATGCCGGCAGGCGCAGCAGCATGTGGAATCCTTCGCGCGT	7356
QY	1770	ACTAGAGCCAGCGGCTCAGCGCTTTAGAAACTACATTTAAACCGGAGGATATAAGTATCCGC	1829
Db	7355	GCTCGAGCGCATCGCGACGSGCTTCCCGCAACTACCAGAAGGCAAGTACACCTGCGCGC	7296
QY	1830	TGAGGAATGCTCTAGACCGGGCGGAGCTTCTGCTGCTTTGCGCACCAAGAAATGACTGC	1889
Db	7295	CGAGGCGCTGCTGGTTCGACAAAGCGGCAATTCCTGACGCTACCGCGCGGAGATGACGT	7236
QY	1890	TTTGGTAGGCGGTATCGGTGTACTTGGGCACCAACTACGACGGTTCGCGAGCATGGAGTGTT	1949
Db	7235	GCTCGTGGGCGGCTGCGCGTGTCTGACACGAAATCAGGCGACACCGCGCACGGTGTCTT	7176
QY	1950	TACAAATTAAGCGGGTCAGCTATCTCAATGACTTCTTTGTAACCTGCTAGACCTCAACAC	2009
Db	7175	CACCCAGCGCGGAATCGCTGACCAACAGACTTCTTCGTGAACCTGCTGCACATGGGCAC	7116
QY	2010	TAAATGGCAGCCAGCGATGAATCAGACAAAGTTTTTTGAAGGCAGAGACTTCAAAACTGG	2069
Db	7115	GGAGTGAAGGCGACGACGACGACGGGAC--GTGTTGAAAGGGCGACCGCGCAACCGG	7059
QY	2070	CGAAGTAAAGTGGAGTGGCACCGCGGTAGACCTGATCTTCGGATCCAAATTCGCGAGTAAG	2129
Db	7058	CGCACTCAAGTGGACCGGACCGCGGGTCTGCTCTTCGGCTCGCACTCGCAGCTCCG	6999
QY	2130	AGCCCTCGCAGAAGTGTACGGCTGTCGAGATTTCTGAAGAAAGTTTCTTAAAGATTTTGT	2189
Db	6998	TGCCCTGGCGAGTCTACGGCAGCGCGGACGCCGCGCAAGTTTCTCGTGACTTCGT	6939
QY	2190	GAAGCGCTGGGCCCAAGTAATGGACCTGGACCGGTTTGATCTGAATATAA	2238
Db	6938	CGCGGCTTGGGACAAAGTGATGAACCTCGACCGGTTTCGATCTCGCGTAA	6890
RESULT 5			
LOCUS	AP003010	340857 bp DNA linear BCT 15-MAY-2001	
DEFINITION	Mesorhizobium loti DNA, complete genome, section 17/21.		
ACCESSION	AP003010	BA000012	
VERSION	AP003010.2	GI:14026358	
KEYWORDS			
SOURCE			
ORGANISM	Mesorhizobium loti		
REFERENCE	Mesorhizobium loti		
AUTHORS	Mesorhizobium loti		
	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;		
	Phyllobacteriaceae; Mesorhizobium.		
	1 (sites)		
	Kaneko,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S.,		
	Watanabe,A., Ideawa,K., Ishikawa,A., Kawashima,K., Kimura,T.,		
	Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsumoto,A.,		
	Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimpo,S., Sugimoto,M.,		
	Takeuchi,C., Yamada,M. and Tabata,S.		
	Complete genome structure of the nitrogen-fixing symbiotic		
	bacterium Mesorhizobium loti		
	DNA Res. 7 (6), 331-338 (2000)		
JOURNAL			

MEDLINE PUBMED REFERENCE AUTHORS TITLE JOURNAL	21082930 11214968 2 (bases 1 to 340857) Kaneko,T. Direct Submission Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research, Yana 1532-3, Kisarazu, Chiba 292-0812, Japan (E-mail:kaneko@kazusa.or.jp, URL:http://www.kazusa.or.jp/rhizobase/, Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934) On May 11, 2001 this sequence version replaced gi:11994985. Location/Qualifiers	gene CDS	<p>/db_xref="GI:14026362" /translation="MTHRADYPLAETOPGAVKGRKSLAEITLDSVLAGDVTMEDL RITPQALQADVARDVGRPTLALNFERGAELVEVQDFIMQVIYELLRPGRAKSEEL LQAATTMRDYOARIAREIEEAETAAARGLFTFR" complement(3601..5871) /gene="ml16722" complement(3601..5871) /gene="ml16722" /codon_start=1 /transl_table=11 /product="glycerol dehydratase large subunit" /protein_id="BAB52960.1" /db_xref="GI:14026363" /translation="MTPKLNKRWKRFADWDERPLRLDKFAAEDPANGFSAPDPKPK GIGIKDRVISLDGYLEDHYDMIDRFIARHIDPEVAPAMALDSATVARLWDMNVP RETLVLAHGMTPAKLAEVVSOLNALEIAFYAKMRKTPGNOAHVTNAKDDPLQLA ADAATAVAFGEDEIETMRVSRNANSNACAVAGGVRGWTILFQCSSEAEELRIAM AGTYSYAEVSVYCTEKSTDDGDPHSAKFAIAAAYASRGVAKRCSGAGSELLMGPH EAKSLYLAEARCLIQRMGMVQQTGGIDGAPLTATIPGGVRELMAENLIIAVMLE CASGNARSTSEIRVGAKILPYLAGSDLICSGMSILKYDNSFNPILINGELEDY LVQRDFEADGGLTLPESRAIELRERAVEAIAAVEELGSLSPSTEDKSSVYVYASGS DTRSLMRDVSFISEAIKERGITVIDVVKALAKRFRFEAENLDDVKRLSGDYLO TSAMLRNGRIVSANDPNYLGPGSGYRVEERLQNDIRVDLQKEVLRSEALHEK DEARHIRNLPANGSAKDDVVIGISPAFGKLKLYRTTAGHRLSEVLGAMLDAIHAR GLKARVVRFRHTDFTFLGSAARLAGSGIGIGIOAKGTAVIHQRDRQPHNNLELFSN APITRLEHYRAGLANAAYALGEMPEPVVVPVVGGEAMSGRYHARVALIYAIETGLTEA GAAPEDVDILTGVKS" complement(6037..7680) /gene="ml16723" complement(6037..7680) /note="hypothetical protein" /codon_start=1 /transl_table=11 /protein_id="BAB52961.1" /db_xref="GI:14026364" /translation="MKRHWLFLGHLWSLSAGLLLTGLTFFAASLTPTLLPRTVYVQGV SGFLAAGYGVIGVFWKLCVLYLELPAPGRRAAKLVAAGVCLLVACIFLWMAAQWNS IRKMLGLDYESAHPKLVGLLALVTALVALARLQFLTIRFVSTVSRFAPRRVSI VGVIVSVALFWSITSGVFRVALHFDADSVROYDALVEPTARTDPDKTGSASLLD WNEIGRGTGFAVGAEGTPEEIHAFNKKALQPIRVVYGLRSADSPKRAKLALNELIR VGGFORSVLVYVPTGSGWIDPAAMESLEYLQNGDVASVAVQYSYLSLWLLVERDY GADTAALRDRVDYWRIPKNSRPLYLHGLSLGAMNSERSTDMDFVVDGDPQGAALW SGPPYGLRWSRSTDRSPGPAWLPFRNGSVIRFMNQDGVAAADPGFVWGPNRIYVL QYASDVPEDFYRDLYREPDWIKPRGPDVSPLELRYPVVTFLQTLTDMAMATTADIG HGVYAPGDYVEGWIEVAVNQWSLQQLAEKRLHGLRP" 7907..8404 /gene="mlr6724" 7907..8404 /gene="mlr6724" /note="hypothetical protein" /codon_start=1 /transl_table=11 /protein_id="BAB52962.1" /db_xref="GI:14026365" /translation="MSDLVIVYPTAKAEVRQRALFQKEYLINISDAVIATKTEA GNVKNQIMNTTAVAGSGFWGLLIGVIFNPVIGVAGSAGALGALTLDVGINDA FMKDLSATLQPGNAALFVLIHGHTADKVLKSAIKFGGGVVLKTSLEDTKEQILREALSQ ASEAA" 9371..10375 /gene="mlr6726" 9371..10375 /gene="mlr6726" /note="hypothetical protein" /codon_start=1 /transl_table=11 /protein_id="BAB52963.1" /db_xref="GI:14026366"</p>
gene	CDS	gene	CDS

source	1..340857 /organism="Mesorhizobium loti" /mol_type="genomic DNA" /strain="MAFF30309" /db_xref="taxon:381" complement(6..770) /gene="ml16717" complement(6..770) /codon_start=1 /transl_table=11 /product="3-ketoacyl-CoA reductase" /protein_id="BAB52956.1" /db_xref="GI:14026359" /translation="MTQSLKRSVVVGTGSKGIGRKIARVFALSGAKVAIVARHPDQA EAAAEIGHAGFGVGVDTSLASIESALAVVAARNGSIDVLCANAGIRFPOARLEDMTS QOWDEVDTNLGTFHAKAAYPKLTSNOGRIVLTSSITGPTGPGWSHYGAVKAG QIGFWRTACIELAKGIYVNAVMPGNIVTEGLVGLGDEYQKTMAAASIPKRLGTVEDI GFAALYFASKEAGYVGTQITIVDGGQILPESLEALA" complement(851..2377) /gene="ml16718" complement(851..2377) /gene="ml16718" /codon_start=1 /transl_table=11 /product="probable amino acid permease" /protein_id="BAB52957.1" /db_xref="GI:14026360" /translation="MAIETARTIDVSKNSINLLHSAVGLAGVLFLAVTGAAPNSA MLGNVPFAGYGIKTYTAAFLMATIVLTIFSIGVAMASRVYSGGCFSPISQGLGR EYMGAGIGSLACYLSEAFSLTGLFAFFGNLWLSHFQINPWI VIALAVIYIAALA YDAVYSARLGLALILEIMLVIFSLGVIFAHGNTSGDSLNFVKVPTVAEQRVG DYAIPAGEAAGVIGFAFWSWGFEMAPNYAESRDPKRIPOSLIISVIGLGLFTFV SNCAVAAVTEADIMVAKPDSGVNFELTPVQFVGWGLYOLMSLLITLSSFCGMAFH NPARVLYSLAREGVLPQAIATHDHHSKPHKASALQSVLAAVVLLYGLAEGDDPS GQAMGLVTLFVLTGTLGLLVLAQVVSALYHWFKNKGGCSMLTFTVPIAPLISLVOLV LVTTLVANLATGGTNGFARSIPYVGLAILIIVGLIWFVILKSANPRAYENIGHMVNEG complement(2477..3175) /gene="ml16720" complement(2477..3175) /gene="ml16720" /note="hypothetical protein" /codon_start=1 /transl_table=11 /protein_id="BAB52958.1" /db_xref="GI:14026361" /translation="MSWKTFRSLYLLGAPEDDPVLVPAQTALIVDQNTYLRKPD RASLSQQRHRYDLWTFFPHERMHGTVIPNTARLLARQNGIECLFARLACTHKDGRD RSLSQQRHNNLLLPKNDAPQSLVAELQAGDEIVYVTKTDSALTGTNLRLHNLG IRNVICCGIFTDQCSTSVSRSLADESYAVIVLECCAAATEDLHRELEIINNIYCHV MSSGELCKIMALK" complement(3185..3604) /gene="ml16721" complement(3185..3604) /gene="ml16721" /codon_start=1 /transl_table=11 /product="glycerol dehydratase small subunit" /protein_id="BAB52959.1"	gene CDS	gene CDS	gene CDS
--------	--	-------------	-------------	-------------

		LAISPFAYMEAFVWFSEFMWVPGAFIGVPIILVVVTVWCAHEQSTWIAITLLSGGRQST	
gene		10712..11839	
		/gene="mlr6727"	
CDS		10712..11839	
		/gene="mlr6727"	
		/codon_start=1	
		/transl_table=11	
		/product="transcriptional regulator"	
		/protein_id="BAB52964.1"	
		/db_xref="GI:14026367"	
		/translation="MANAPAGWNAHAILEPPANPAEPTLVVPEOTIPCFEESTGD LPPKDPFAARQSFATMLDFVEPGDTEAGFAGTQVINDGLIALAQVSTNGLDFTSLA GHVRDPVDHLYLIQDGRSQITPPSKSFGDAGSVQVHPLGRVFGESVYDSEMLLI FVPRDFCDMSGLAAEFSLRGMGPPLADYLAASVLRPLTHVNDPLGLLSATRA MLQVCIAPSDHLTEADHPISATLLERARVPHLFDPLDLSIQMTRELGTSRRLY RLFPASGIVHYIQHRRLLAAALADPNDRRLIIDIADEYFGDGAFFSNAFRREFG YCPSEVRTGVKNGSPSHMDEVETAAAPGLRLQLLRLLAQ"	
gene		complement(11950..12498)	
		/gene="ml16729"	
CDS		complement(11950..12498)	
		/gene="ml16729"	
		/note="hypothetical protein"	
		/codon_start=1	
		/transl_table=11	
		/protein_id="BAB52965.1"	
		/db_xref="GI:14026368"	
		/translation="MTAYENGTVSSLPPTSVRILLGLVLIYGGIFVLGDLALATRVISA IVIGVTAIVGAFEMHAFNWKHMSGFTWQFLGLTVLVFGMAVAVRQVPSGALVITYV LGLLVASGVVRLVFGVVRHNSDVGMVLSGTGFLAGLVLLTGWPSTGLVWLGLLIG IDLISHGMWLTGWLPTAKTA"	
gene		complement(12495..15662)	
		/gene="ml16730"	
CDS		complement(12495..15662)	
		/gene="ml16730"	
		/codon_start=1	
Query Match 44.3%; Score 991.8; DB 1; Length 340857;			
Best Local Similarity 66.8%; Pred. No. 1.2e-256;			
Matches 1429; Conservative 0; Mismatches 707; Indels 3; Gaps 1;			
QY	90	TGCAGGTGGCGGACCAAAAAGAGGATTTGGTGGCCCAACATGCTCAACCTCGGCATCTT	149
Db	197392	TGGATCGCGCGTGGACCAACCGCTGATTTGGTGGCCGACAGCTGAACGTCAGATCCT	197451
QY	150	ACGCCACATTCATCGCTATCGGACCAACAGCCCGGATTTTGACTATGCCGAAGATT	209
Db	197452	GCATCAGCAGTCCTCTGTGCGGACCGGATGGTGGCCGACAGCTGACCTATGCGGAGG	197511
QY	210	TAAGAAGCTAGATCTGGCAGCGGTTAAAGAGGACTGGCAGCGCTAATGACAGATTCACA	269
Db	197512	CAAGAGCCTCGACCTCGACCGCGTATCAGAGACCTGCACGCGCTGATGACGATTCGCA	197571
QY	270	GGACTGTGGCCAGCAGATACGGTCAATATGGCCCTTCTTTATACGCATGGCGTGCCA	329
Db	197572	GGAGTGTGGCGCGCGGATTTGGCCCACTACGGCCCGCTGTTCATCCGATGGCCTGGCA	197631
QY	330	CAGCGCGGACCTACCGTATCGGTGATGGCGGTGGTGGCGGTGGCTCCGGCTCACAGCG	389
Db	197632	CAGCGAGGACCTTACCGCATTTGCGGACCGCGCGCGCGCGCGGCTGGCCAGCAGCG	197691
QY	390	CTTCGCCCTCTCAATAGCTGGCCAGACAAATGCCAATCTGGATAAGCAGCTTGCTTCT	449
Db	197692	TTTCGCCCGCTCAACAGCTGGCGGACAATGTCAACCTGCAGAAAGCCCGCGGCTTTT	197751
QY	450	TTGGCCCATCAACAAAATACGGTCGAAAATATCTCTGGCGGATCTAATGATCTCAC	509
Db	197752	GTGGCCGATCAAGCAGAAATATGGCGCAAGATCTCTTGGCGGACCTTCTGATCTCTAC	197811
QY	510	AGGAACGTAAGCTGTGGAACACTATGGCTTTTAAACCTTTTGGTTTTCAGGTGGCAGAGC	569
Db	197812	CGGCAACGTCGCGCTGGAATCGATGGGCTTCAAGACCTTTGGGCTTTTGGCGGCGCGCG	197871

Qy	570	AGATGTATGGGAGCCTGAAGAAAGATGTATACTGGGAGCAGAAACGAATGGCTGGGAGA 629
Db	197872	CGACGCTGGGAGCCTGAGCAGACGTTTACTTGGGTCCCGAAGCAAGTGGCTGGCCGA 197931
Qy	630	CAAGCGCTATGAAGTGAACCGAGAGCTCGAAATCCCTTGGAGCGGTACAATGGGACT 689
Db	197932	CGAGCGCTACAGCGGACCGGTGACCTCAGAAACCGCTGGCGCGGTGCAGATGGGCT 197991
Qy	690	CATCTATGTAAACCCCGAAGGACCAAGGACCAAGGACCAAGCCCTATCGTGTGCGCGTCA 749
Db	197992	GATCTACGTCAATCCGAGGAGCCAGCGCAATCCGATCCGCTGGCGCGCGCGGCA 198051
Qy	750	TATTCGTGAGACTTTTGGCCGAAATGCAATGAATGACGAAGAAACCGTGGCTCTCATAGC 809
Db	198052	TATCAGGACACATTCGCGCGTATGGCGATGAACGACGAGAAACCGTAGCACTCATCGC 198111
Qy	810	GGGTGACACACCTTCGGAANAACCCATGGTGTCTCCCATGCGGAGAAATATGTGGCGCG 869
Db	198112	CGGCGGCATACGTTTCGGAAGACCCATGGTGGGGGTGACGC--GAGCCTGGTGGGTGT 198168
Qy	870	AGAGCTCGCGCGGAGCTATTGAAGAAATGAGCCTGGGTGGGAAACACACTACGCGCAC 929
Db	198169	CGAGCGGAGCGCTGATATCGACGACAGGGCTTGGCTGGGCGACCAATTCGCGCAC 198228
Qy	930	CGGACCGGTGGGATACCATCACAGTGGACTAGAGGCGCCCTGGACCAAGACCCCTAC 989
Db	198229	CGGCAAGGCGGTGACGCCATCGGACGCGGTCTGGAAGTCATTTGGACGACGACCGCAC 198288
Qy	990	TCAATGGAGCAATACCTTTTGAACCTCTTGGTTTACAGTGGGAGCTTACCAAAAG 1049
Db	198289	CAATGGAGCAACACTTCTTCGACAACTCTTGGGTTCGACTGGGAACCTGACCAAGAG 198348
Qy	1050	TCCAGCTGGAGCTTATCAGTGGAAACCAAAAGAGGCTGCGGGGTGGCACCATACCGCA 1109
Db	198349	CCGCGAGCGCGCATCAATGACCGCGAAGGTTGGCCCGGTGCCGCGACGGTGGCGGA 198408
Qy	1110	TGCACATGATCCGACAAAGTCGACGCTTCATTTATGTCTACTACGGACCTGGCGCTGCG 1169
Db	198409	CGCACAACTCGGCAAGCGCTCACGCCCTCCATGCTGACCAACCGACCTCGCGCTGCG 198468
Qy	1170	CATGGACCTGATTACGAAAAAATTTCTGACGCTTATGAAACCTGATGATTTGCT 1229
Db	198469	TTTCGATCCGCTTCGCAACGATCTCGAACGGTTCATGGAATCCGGATCAGTTGCG 198528
Qy	1230	AGATGCTTTCGGAAGAGTGTACAACTGACACACAGATATGGACCAAAAGTGGCG 1289
Db	198529	CGAGCGCTTCGCGCGCTGTTACAAAGTCAACCGACCGGACATGGCCCGGTGCGCGG 198588
Qy	1290	CTACCTGGACCAAGTGTCTCAGGAGACCTCATCTGGCAACCGCTATACCAATGT 1349
Db	198589	TTATCTCGGCGCGTTGGTTCCGAAGGAGAGCTGCCCTGGCAGGACGTCATCCCGGTGT 198648
Qy	1350	AAGCATCTCTTGTAGACGAAAAACGATATTGAAGGCTTAAAGCCAAAATCTCTGGAATC 1409
Db	198649	CGATCATGTGTCATCGACGAGGATGCGGAAGCGCTCAAGGCTGAGATTTGGCGCTC 198708
Qy	1410	GGGACTACGGTAAGCAGCTGTGTAAAGCAGGCTGGGCTTCTGCATCTACTTTTAGAAA 1469
Db	198709	CGGCGTGTGCTGTCGACGTGTTTCGACGGGTGGGCTCGGCTTCCACCTTCCCGCGG 198768
Qy	1470	CTCTGACAAAGCGGCGGTGCCAACGGTCACGTATAGAGTGGCGCCCAAAAGACTG 1529
Db	198769	TTCTGACAAAGCGTGGCGGTGCCAATGCGCGCGCATCCGCTCAGCCCGCAGAGGACTG 198828
Qy	1530	GGAGTAAACAAACCTCAGCAACTTGGCAGGCTACTCAAAACACTAGAAGTATCCAGGA 1589
Db	198829	GGCGTCAACACCGCGCGGAGCTCGCCAAAGGTGCTGCAGAACTCCAGCGCATTCGCCA 198888
Qy	1590	GGACTTTAACACAGCGCAATCAGATAACAAAGCAGATATCTTGGCCCACTGATTTGCT 1649
Db	198889	GGACTTCAATGCCGACAGACCGGACCAAGAGGTCTCGCTTGGCGATCTGATCTGTTCT 198948
Qy	1650	GGCGGCTGTGGGTGTAGAAAAGCTGCAAAAGATGCTGGCCATGAGGTGCGAGTGC 1709

Db	18949	
----	-------	--


```

/translaton="MDOKSDSAGKCPVAHTAPGRSNNRQWPDQDVLQVLIHRHSGLSLSD
PLNTFNFAEEFKDLDDLDLRLALMTSDQWMPADFGHYGSLFIRMAWHASGTYR
ITDGRGAGQORFAPLNSWPONALNDKARLLIPIKQYGNRISWADLLILTNVA
LESQKFTFGAGRVVDVWEPELFWGPEGTWLDGDRYSGEROLSEPLAAVOMGLIIV
NPEPGNPDVPAARDIRTEFARMANDETVALLAGHTFGKTHAGDPSFELGADP
EGNAIEDQGLWSTGTGKDAITGCGPVTWSQTPTRSNHFFENLFWHEWELTKS
PAGAYQWKNANAKATIPADYDPSRKHVPTRLTDLISLRFDPAYEKISRRLENDPEFA
DAFARWFLTHRDMGPKVRYLGLPEVPAEDLIWODVIPAYDHRLEVDTEIAGLKAKII
ASGLSVQELVSTAWASASTPRGSDKGANGARIRLAPQKDWVNRPAQLARLSVLE
GIORDFNAOQDKKISLADLIIVLAGGAAGAKAAGGCHDITVPFTPCRMADASEAOTD
AASFALEPRADGPRNVSTTROQFMKPEALVDRAOLLTLTAPENTVLVGLRLVKA
GEKHGVETSPRALNDFVFNLLDGTQSPTEGEGVYEGEDRRRTGAARWGTGRVD
LIGSHSQLRAFAEVAQAQDAREKFVKDFVAAMTKVMNADRFDLV"
5640. .6113
/gene="Sma2381"
CDS
5640. .6113
/gene="Sma2381"
/function="Miscellaneous; Unknown"
/notes="glimmer prediction"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK65945.1"
/db_xref="GI:14524461"
/translaton="MLLKAMKIGRSQSLYDTFGDKWLYCLAVERYSASETSAAHIALLR
GKPRALLEGIRLWVRVVDNANQACLVNSICEFGQSRPDIAALHRAADRRLKNVYER
IREQAAGDIALVSLPAEAVADFLVANIAGIRIARAAGCARREHLQSLSLRALRAVT"
6167. .7135
/gene="Sma2383"
CDS
6167. .7135
/gene="Sma2383"
/function="Small molecule metabolism"
/notes="glimmer prediction; similar to TED2, a putative
quinone oxidoreductase, from Zinnia elegans, BAA06460, and
from cowpea, T11672; also similar to putative
oxidoreductases from Mycobacterium tuberculosis,
Agrobacterium tumefaciens, etc. Zinc-binding family"
/codon_start=1
/transl_table=11
/product="probable oxidoreductase"
/protein_id="AAK65946.1"
/db_xref="GI:14524462"
/translaton="MAVYMKYVGVDVDFVDRPVPARVHVVEVAAGVNFMDI
GVQGMATWDIPKVLGVGEGARVLAVGDGTEFAGVDRVAVWYAPGSYAQRQSTPA
ASLVKIPDVTDDRTASTMOGLTASHFATDFYPQGDIALVHAAGVGLLITQII
RLRGVRVIGRSSDVAIARKAGABSHFIVDTDRFADEVLRITGEGVNVYDGGSP
KTFKSGLEALRRSGTCWYGPVLGGPCLPIMNPKSIKIGYATFMDHIHTRELLDR
TKOLFWDIEPDSITVIGETRYRLADAAGAHAASRATTGKLLIP"
complement(7357. .8946)
/gene="Sma2385"
CDS
complement(7357. .8946)
/gene="Sma2385"
/function="Cell processes; Transport of small molecules"
/notes="glimmer prediction; similar to probable ABC
transporter from Streptomyces coelicolor, T35245, and to
varM (probable transporter for virginiamycin) from
Streptomyces virginiae, BAA96297. ABC transporter family"
/codon_start=1
/transl_table=11
/product="probable ABC transporter, ATP-binding protein"
/protein_id="AAK65947.1"
/db_xref="GI:14524463"
/translaton="MPSITLSALSWSKPDGEHVFSDILAFGPRTGLVRNGRIGKTS
VLNIIAGTLPSSGTVAIQARVALARQILRAGADETADVFQATQAVAVLRAEKGDIA
SVLELETADWTVRIVSALARGLERADTLLNQLSGGORTRAVLAAALFSPDFLL
LDPTNLRDGRRAVYLLGLSGWRSGNIVVSHDRELLEMDAIIELTSLGTRKYGSGG
SAYQARAVELEAAQSLILARTADEVDKARALARLERLDKRDASTRKAAGDMFRI
LVGRKSNESRGKSVELAEARRAGDALDAATAKARIEVLQPFSLRPLTELPAQRQ
VLAFDGTAGYDAPRIIRLDSFVPRVSVTPGNCGKTSLLKVLVTELPFGFT
VSNVPTLIDQSVSLTERGETILENFKRLNPGASDNACRAALASFRFRADALQRVE
ALSGGOVLRAGLACALGGSDPPSLILLIDPTNHLIDSTEAVEAGLLSYDGLVYVSH
DETFLANIGTIRVELSTRG"
9057. .9713

```

```

/gene="Sma2387"
9057. .9713
/gene="Sma2387"
/function="Miscellaneous; Unknown"
/notes="glimmer prediction; putative bacterial regulatory
protein, tetR family"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK65948.1"
/db_xref="GI:14524464"
/translaton="MIRARRAGRPRTREAEALTRLLDSARSFTFARKGTANSMBEE
IAAPLGSKHTLYRNPQRQALLEAVVRLVFRKTLAEAGGCGEAPLALRDMAR
YFRFGTRDYSAFVLSVTAERLPLRERLAWSAALPLVQALISAOAGLVVPG
STIEICHVLIDLEGANNRVRLCLSESPPDASERLRLFESRWAVFQTAMMPEPNRLPGV"
complement(9222. .10455)
/gene="Sma2389"
CDS
complement(9222. .10455)
/gene="Sma2389"
/function="Cell processes; Adaptation"
/notes="glimmer prediction; very similar to Ohr, a putative
stress-induced protein, from Bradyrhizobium japonicum,
Query Match 42.9%; Score 960; DB 1; Length 10578;
Best Local Similarity 67.1%; Pred. No. 4.2e-248;
Matches 1445; Conservative 0; Mismatches 685; Indels 24; Gaps 5;
QY 85 CAAGTGCAGGTGGCGGCACCAAAAACAGAGGATTGGTGGCCCAACATGCTCAACCTGGC 144
Db 5065 CATACGCCAGCAGAGGAGCAGATCGAACCGCGACTGGTGGCGGATAGTTGGACGTCCAG 5006
QY 145 ATCTTACGCCAACATTCATCGCTATCGGACCCAAACACCGCGGATTTTGACTATCGCGAA 204
Db 5005 GTTCTGACCGGCATTCGGCCCTTCGATCGCGCTGGGCAATAGTTCACTAGCCGAG 4946
QY 205 GAGTTTAAAGCTAGATCTGGCAGCGGTTAAAGAGGACCTGGCAGCGCTAAATGACAGAT 264
Db 4945 GAGTTCAAGAAGCTCGACCTCGATCGCTGAAGGGGACCTGCGTGACATGATCAGCAT 4886
QY 265 TCACAGACATGGTGGCCAGCAGATTACGGTCAATTATGGCCCTTCITTTATACGATGGCG 324
Db 4885 TCGCAGACATGGTGGCGCGGAGCTTCGGCCACTATGGCGGGCTCTTCATCCCGATGGCC 4836
QY 325 TGGCACAGCGCCGACCTACCGTATCGGTGATGGCGCTGGTGGCGGTGGCTCCGGCTCA 384
Db 4825 TGGCACAGCGCTGCACCTACCGTATCACCGATGGCGCTGGCGCGCGCCGACGGTCA 4766
QY 385 CAGCGCTTGGCGCTCTCAATAGCTGGCCAGACAAATGCCAATCTGGATAAAGACAGCTTG 444
Db 4765 CAGCGTTTCGCGCGCTCAATAGCTGGCCGACAAACGCCAATCTCGACAAGGCCGACGT 4706
QY 445 CTTCCTTTGGCCCATCAACAAAATACGGTCGAAAATCTCTGGCGGATCTAATGATA 504
Db 4705 TTGCTGTGGCGGATCAACACAGAAATACGGCAACAGGATTTCTCTGGCGGATCTAATGATT 4646
QY 505 CTCACAGAAACCTAGCTCTGGAACATATGGCTTTTAAACTTTTGGCTTTTTCAGGTGGC 564
Db 4645 CTGACCGCAATGTGCGGCTCGAATCCATGGCTTCAAGACCTTCGGTTTTCGCGGGGC 4586
QY 565 AGAGCAGATGTATGGGAGCCTGAAGAAGATGTATATCTGGGAGCAGAAACCGAATGCTG 624
Db 4585 CGSGTGACATATGGGAACCCGAGGAA--CTGTTCCTGGGACCGGAAGCACCCTGGCTG 4529
QY 625 GGACAGAGCGCTATGAAGCTGACCGAGAGCTCGAAAATCCCTTGGGAGCGCTACAATG 684
Db 4528 GGGCAGGCGCTACACGGCGGACGCAACTGAGCGAGCGCTTGGCGCGGTGAGATG 4469
QY 685 GGACTCATCTATGTAAACCCGAGGACCCAAACCGCAAGCAGACACCTCATCTGCTGCTG 744
Db 4468 GGTCTCATCTATGTAAACCCGAGGACCCGCAACCGCAACCCGACCCGCTGCTGCTGCG 4409
QY 745 CGTGATATTCGTGAGACTTTTGGCCGAATGGCAATGAATGACGAAGAACCGGTGGCTCTC 804

```


Db	4408	CGGATATCGGGAACCTTTGACGCATGGCGATGAACGACGAGAGACCCTGGGCGTG	4349	3337	TTCATGAAGCCGGAAGAGGCTCTGGTGGATAGGGCTACGCTGCTGACGCTGACCGACCG	3278
QY	805	ATAGCGGGTGGACACACCTTCGGAAAAACCCATGTGTGCTGCGCATCGCGAGAAATATGTG	864	1879	GAATGACTGCTTTGGTAGCGGTATGCGTGTACTGTGGCACCACCACTACGACGTTGCGCAG	1938
Db	4348	ATCGCGGTGGCCACACCTTCGGCAAAACGCATGGCGCGCATCGCTCG--TTCATC	4292	3277	GAATGACCGCTTCTCGTCGCGGCTCGGGTGTG-----AAGCGGGCGAGCCCAAG	3224
QY	865	GGCCGAGAGCCCTGCGCCGAGGTATTGAAGAAATGAGCCTGGGGTGGAAAAACACCTAC	924	1939	CATGGAGTGTTCACAAATAAAGCGGGTCACTATCCAAATGACTTCTTTGTAACCTGCTA	1998
Db	4291	GGTGCAACCCCGGAAGCGCGCCCATFAGAGGATAGGGTCTTGGCTGGAAGACGACCTTC	4232	3223	CACGGCGTGTTCACGTCCCGCCCGAAGCGCTGACGAAGACTTCTTCGTCACCTGCTC	3164
QY	925	GGCACCGGACACGGTGGGATACCATCACAGTGGACTAGAAAGCGCCCTGGACCAAGACC	984	1999	GACCTCAACACTAAATGGCGAGCCAGCGATGAATCAGACAAAGTCTTTTGAAGSCAGACAG	2058
Db	4231	GGCACTGGCGTGGCAAGGACCGGATCACGGCGGCGCGAAGTGACCTGGTTCGAGACC	4172	3163	GATATGGGCACGCGATGGTCAACCCATCGAAGGCGGAAGAGCGGTATACGAAGCCCGGAC	3104
QY	985	CTTACTCAATGAGCAATTAACCTTTTGTGAAACCTCTTTGGTTACGAGTGGAGCTTACC	1044	2059	TTCAAACTGGCGAAGTAAAGTGGAGTGCACCCGCGGTAGACTGATCTTCGATCCAAT	2118
Db	4171	CGACCCGTTGGAGCAACCACTTTTTCGAAAAACCTGTTCAACACCAAGTGGAACTGACC	4112	3103	CGCAGACGGGCGCTGCCAGTGGACCGCAGCCGAGTGCACCTGATCTTCGGCTCGCAT	3044
QY	1045	AAAAGTCCAGCTGGAGCTTATCAGTGGAAACCAAAAGACGGTCCCGGGCTGGCACCAT	1104	2119	TCCGAGCTAAGAGCCCTCGCAGAAAGTGTACGGCTGTGCAGATCTCTCAAGAAAGTTGTT	2178
Db	4111	AAGAGCCCGCGCGCGCTATCAATGGAAGGGGAAG-----CGCCGAGGCCACCATC	4058	3043	TCCGAGCTCGCGCGCTTTCGGAGGTGTACGCCCAATCCGATCCGAGGAGAGTTTCGTG	2984
QY	1105	CGGATGCAATGATCCCAAGCAAGTCGACGCTCCATTTATGCTCACTACGACCTGGCG	1164	2179	AAAGATTTTGTGAAGCCCTGGCGCAAGTAATGGACCTGGACCGGTTTGTATCTG	2232
Db	4057	CGGATGCTATGATCCCTCCAGGAAGCATGTTCCGACAGGCTGACACGACCTGTCTG	3998	2983	AGGACTTCTGTGCGGCTGGACCAAGTCATGATGCCGCGCTTGTATCTG	2930
QY	1165	CTGCGCATGACCCCTGATTACGAAAAAATTTCTCGACGGTACTATGAAAACCCCTGATGAG	1224	RESULT 7		
Db	3997	CTTCGTTTCGATCCCGCTTACGAGAAATCTCGCGCGCTTCTCGAANAATCCGATGAG	3938	AE016872		
QY	1225	TTTGCAGATGCTTTCGGAAGACGTGTTACAACTGACACAGAGATATGGACCAAG	1284	LOCUS	AE016872	AE016853
Db	3937	TTGCGGAGCGCTTCCCGCGCGCTCGTTCAAGTGCACCGCACATCGGCGCGAAG	3878	DEFINITION	Pseudomonas syringae pv. tomato str. DC3000 section 17 of 21 of the complete genome.	
QY	1285	GTGCGCTACTGGGACGAGAGTCCCTCAGGAGACCTCATCTGCGCAAGACCTATACCA	1344	ACCESSION	AE016872	AE016853
Db	3877	GTTCGCTATCTCGGGCTGAAGTCCCGCGGAGGACCTGATCTGGAAGACGTGATCCCG	3818	VERSION	AE016872.1	GI:28854817
QY	1345	GATGTAAAGCATCTCTTGTAGCAAAAAACGATATTGAAGCGCTAAAGCCAAAACTCTG	1404	KEYWORDS	Pseudomonas syringae pv. tomato str. DC3000	
Db	3817	GCTGTGACACACAGCTGTGTCGACGAGACGGACATAGCCGCTCAAGCGGAAGTTATC	3758	SOURCE	Pseudomonas syringae pv. tomato str. DC3000	
QY	1405	GAATCGGACTGACGTAGCAGCTGTGTAAGCAGCGATGGCTCTCTGCATCTACTTTT	1464	ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;	
Db	3757	GCTTCGGGCTGTGTCGAGGAACCTGTCTCGACCGCTTGGGCTTCCGCTCGACCTTC	3698	REFERENCE	1 (bases 1 to 311963)	
QY	1465	AGAACTCTGACAAAGCGCGGTGCAACGGTGCACGTATACGACTGGCCCGCCACAAAA	1524	AUTHORS	Buell, R., Joardar, V., Khouri, H., Fedorova, N., Tran, B., Russell, D.,	
Db	3697	CGCGGCTCGACAAAGCGCGCGCCCATGTGCGCGCATCGCTGCGCGCGCAGAA	3638		Berry, K., Utterback, T., Van Aken, S., Feldblyum, T., Gwin, M.,	
QY	1525	GACTGGAAAGTAAACAACCCCTCAGCAACTTGGCAGGTACTCAAAACACTAGAAGTATC	1584		Dodson, R., Deboy, R., Durkin, A., Kolonay, J., Madupu, R.,	
Db	3637	GACTGGAGGTCAACCGCGGCTCAGCTCGCCAGAGTTCTCTCTGCTTGAAGTATA	3578		Daugherty, S., Brinkac, L., Beanan, M., Haft, D., Selengut, J.,	
QY	1585	CAGGAGACTTTAACCAGCGGCATCAGATAACAAGCATATCTGTTGCCGACCTGAT	1644		Nelson, W., Davidlson, T., White, O., Fraser, C. and Collmer, A.	
Db	3577	CAGAGGACTTCAACCGCCCGCAGACCGCGCAAGAAGATCTCGCTTCCGACCTGATC	3518	TITLE	Complete Sequence of Pseudomonas syringae	
QY	1645	GTGCTGGCGGCTGTGCGGTGTAGAAAAAGCTGCAAAAGATGCTGGCCATGAGTGCAG	1704	JOURNAL	Unpublished	
Db	3517	GTGCTTGGCGGCGCGCTCGAGAGGCTGCCAAGCGCGCGCGCCACGACATCAC	3458	AUTHORS	2 (bases 1 to 311963)	
QY	1705	GTGCTTTCAACCCCGGAGCGGATGCCCGCTGAGCAACCGATGTGGAAGCTTTC	1764		Buell, R., Joardar, V., Khouri, H., Fedorova, N., Tran, B., Russell, D.,	
Db	3457	GTGCGCTTACACCGGCGCGATGGATGCTCGGAGGCAACAACCGCGCATCTTC	3398		Berry, K., Utterback, T., Van Aken, S., Feldblyum, T., Gwin, M.,	
QY	1765	GAAGCACTAGACCGCGCTGACGCTTTAGAACTATTAACCCGGAGCAATAAGTA	1824		Dodson, R., Deboy, R., Durkin, A., Kolonay, J., Madupu, R.,	
Db	3397	CGGCGCTGGAACCTCTCGCCGACGGTTTCCGCAACTATGTGAGCACCAACCGCTCAGCA	3338		Daugherty, S., Brinkac, L., Beanan, M., Haft, D., Selengut, J.,	
QY	1825	TCG-----CTGAGGAATGCTCTAGACCGCGCGCTCTGCTTTCGGCACCA	1878		Nelson, W., Davidlson, T., White, O., Fraser, C. and Collmer, A.	
					Direct Submission	
					Submitted (03-MAR-2003) The Institute for Genomic Research, 9712	
					Medical center Dr, Rockville, MD 20850, USA	
					Location/Qualifiers	
					1. 311963	
					/organism="Pseudomonas syringae pv. tomato str. DC3000"	
					/mol_type="genomic DNA"	
					/strain="DC3000"	
					/db_xref="taxon:232283"	
					/note="pathovar: tomato"	
					104. 547	
					/locus_tag="PSPT04431"	
					104. 547	
					/locus_tag="PSPT04431"	
					/note="similar to gp:9950675; identified by sequence	
					similarity; putative"	
					/codon_start=1	
					/transl_table=11	
					/product="conserved hypothetical protein"	
					/protein_id="AA057880.1"	

```

/db_xref="GI:28854818"
/translation="MEHSLVLLPTLALVAGVYIGFLVARLLPNTVPNSQRLDDI
QERDFTQNEVYTFNSANLVQKLSQYQDVQSHLAEGANRLALDQLTRQLLAALH
PEANQHRDRDLTPPRDTPTKDYAPKSPNTPGMLDHYGLKK"
complement(677..2575)
/gene="cysN/C"
/locus_tag="PSPT04432"
complement(677..2575)
/gene="cysN/C"
/locus_tag="PSPT04432"
/notes="similar to SP:050274, GB:233105, and PID:530438;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="sulfate adenylylate transferase, subunit
1/adenylylsulfate kinase"
/protein_id="AAO57881.1"
/db_xref="GI:28854819"
/translation="MHSQSDLSIEDILAYLGQHERKEMRLRFLTCGNVDGKSTIGRL
LHDSKIYEDHLEAITRDKSKSGTGDVDLALLVDGLQAREQGITIDVARYFSTA
RKFTIADTPGHEQYTRNMGASTCOLAILVDARYGVQOTRRHSYIASLLGIKHI
VVAJNKDNGFQDERVFESIKADYLKFAEGIAFKPTTMAFVPMsALKGDNVVKRS
PYAGOSIMELETFELASDRNYTDLPBPQVYVNRPNLNFEGFAGTLASGIYHKGDEI
VLPDSKSRVKSIVTEGELEQAGPGQAVTLTMEDEIDISRGDLLVHADNPQVSDA
FDALVMAEPPMLPGKKYDKIKRATSYVPGSIATIHVDVNTLEEPPASSLQINEIG
RVKYSLDAPIALDYSNRTTGAFFIVIDRLTNGTVAAGMIITAKPVSGSGSHHGELAH
VSTBERAQRFQQPATVLFSLGSGAKSTLAYAVERKLFQMGRAVYVLDGQNLRHDLN
KGLPDQAGRTENNRAAHVARQFNEAGLLTLAAFVAPDAGEERAKALIGKERLTV
VVOASPAVCRERDPQGLYAADGDNIPGESFPYDPLNADLVDTQTLSVESRVQVLD
LLRGKAI"
complement(2586..3503)
/gene="cysD"
/locus_tag="PSPT04433"
complement(2586..3503)
/gene="cysD"
/locus_tag="PSPT04433"
/notes="similar to GP:11992146, and SP:P21156; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="sulfate adenylylate transferase, subunit 2"
/protein_id="AAO57882.1"
/db_xref="GI:28854820"
/translation="MVDKLTHLKQLEAESIHIREVAAEFDPNPMVLSIGKDSAVMLH
LAKAFPGKLPFPMVMDVTRKQFQEMRFRDQMVEMGLDLITHINPDGVAQINFP
THGSAKHTDINKTEGLKQALDKHGDAFGAGDERKSAKERVYSFRDSKHRWDPK
NORPELVNNGVNGKESIRVPFLSNKNTLDIWQYIYLEGIPVPLVFAAERDVIK
NGTLIMIDDERLEHLDEEKSRIYVKKVFRFTLGCYPLTGAVESEATSLTDIIQEML
LTRFSERQGRVIDHDGAGSMEEKRQGF"
complement(3657..4415)
/locus_tag="PSPT04434"
complement(3657..4415)
/locus_tag="PSPT04434"
/notes="similar to GB:X16150, GB:M32406, GB:M32407,
GB:M32408, GB:M32409, GB:M32410, GB:M32411, GB:M32412,
GB:M32413, GB:M32414, SP:PI4151, PID:307134, PID:34425,
PID:38093, and PID:386860; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein TIGR00486"
/protein_id="AAO57883.1"
/db_xref="GI:28854821"
/translation="MAVALSTLVEADRYLNSARIODYCPNGLQVEGRPOVMRTVSGV
TASQALLDAVEAQADILVHHGYFWKGENPCVGMKORLTKLKHIDISLLAYHLPL
DVHPEVNNVQLARQLDITVEGDLDPENARVUGLIGLSEPMTHARDARRVQDQALGRE
PLLEGGQMLRRVGCWCTGGGGYIDQAVLEGVLDLFGSEASEQTFFHSARENDISFIAA
GHATRYQVQALGDYLCARRFALEHIFIDCPNPI"
4527..5687
/locus_tag="PSPT04435"
4527..5687
/locus_tag="PSPT04435"
/notes="similar to GB:231695, and PID:469144; identified by

```

```

sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="trypsin domain protein"
/protein_id="AAO57884.1"
/db_xref="GI:28854822"
/translation="MFKALREFGWPPLAGVLIALLIITQRYPOWVGLPSMDVNLQAAPQ
TTVMQGPSSYADAVIAAFAVNVLYTTKVNKGTNPLFDPQFRFRFGDNTPKQRM
ESSLGGVMSPEGVILITNNHVTGADIVVALKDGRETIARVIGNDPTDLAVLKID
LKNIPATILASDGIIRGDVALAIGNPFGVCQOTVMGLIISATGRNQLGNTYEFIQT
DAALNPGSGALVDAAAGNLGINTAIFRSKGGSGGIGFAIPTKLAMDVMKSIIEHQQ
VIRWGLIEYQPLTPELAESGLKDRFGIVVGFDRGPAQKAGRLGDLVILANGEP
AGDGRSMNQVATPKPKDKIAIDVMRNGKENRLSAEVLGRPPPTAAAPE"
complement(5712..6164)
/locus_tag="PSPT04436"
complement(5712..6164)
/locus_tag="PSPT04436"
/notes="similar to GP:15979604; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAO57885.1"
/db_xref="GI:28854823"
/translation="MQSNTQPNPAPFLKAGETVVFQVCKLGNWAKFIIRHRRQ
RIRLTVQSPGEGALLWAGLPDRFDTMAAVTGDRLVYRSEAFVALSDPPAPRWL
RVLATFPVAVIRDLWYDRLALNRYRLFGRYDSCLLPTPDHRRFLKSP"
complement(6921..7973)
/gene="hisC"
/locus_tag="PSPT04437"
complement(6921..7973)
/gene="hisC"
/locus_tag="PSPT04437"
/notes="identified by match to TIGR protein family HMM
TIGR01265"
/codon_start=1
/transl_table=11
/product="histidinol-phosphate aminotransferase"
/protein_id="AAO57886.1"
/db_xref="GI:28854824"
/translation="MSKFWSPFVSGLYPVVPGEQPKLTRLVKLTNNPNYCPSPKAVE
AMRAALDDLRLYPDPNSDLLKHAVAGTVYVQNVQVFLGNSDVLAHIFHALFQHDA
PLLEPDISYFSPYVYGLYIDYEMVALDEQFIARADYARNAGIIFPNPNAPTGL
LGLDAVQSIITKASPSVYVDEAYIDFGETAITLVDRYPNLLVTQTLKSRSIAGLR
VLAVGHPDITLALERNKSNFSYPLDRMANVGGAAAFEDREHETFRNKVIESREAL
VEOLRGSGDFVLPSSAANFIFARHPQHDAALAAKLRQGVIVRHFQOORINQFLRISI
GTPQHOALLGLSDI"
complement(7973..9319)
/gene="hisD"
/locus_tag="PSPT04438"
complement(7973..9319)
/gene="hisD"
/locus_tag="PSPT04438"
/notes="similar to SP:034651; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="histidinol dehydrogenase"
/protein_id="AAO57887.1"
/db_xref="GI:28854825"
/translation="MTPTAIRRLDAADPDFARHLHLLSWESDSDSVNQVRVLDIIR
AVRERDAALVELTKDFGLQVASMELILPRERLELATRITPVQALKEAERVY
SYHKQKQDSWSTYQADGTVLGOKVTPLDRAGLVVPCKKASYPSSVLNKAIPKAVGV
TEVMVYPTRGEINELYLAAACIAGVDVFTTIGCAQAAALAYGTSVSPKVKVGP
GNITYATKRVFGVGDIMTAGSEILVVCDDGTDPDITAMDILFSQAEHDEDAQA IL
VSPVAFEDLKVAASITRLLPTMARAAIVETISINGRALKIVADMAQAIERANRATPEH
LELSVADPEAWLPHIRHAGATFMGRHTSEALGYCAGNHVLPISGTAFFSSPGLGVD
FOKRSSIIYCSPEGASLGKTSVLARGESLSGHARSAEYRITLTDWKTNLEBEGK"
complement(9444..10079)
/gene="hisG"
/locus_tag="PSPT04439"
complement(9444..10079)
/gene="hisG"

```

```

/locus_tag="PSPT04439"
/note="similar to SP:034520; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ATP phosphoribosyltransferase"
/protein_id="AA057888.1"
/db_xref="GI:28854826"
/translation="MTIALSKGRILDDPLPILABAGIVPTNPDKSRKLIIPPTQAD
VRLILVRATDPTVYVYHGAADLGVAGKDLMEYTGGLYELDLQIAKRLMTAGAI
Query Match      42.5%; Score 951.2; DB 1; Length 311963;
Best Local Similarity 65.6%; Pred. No. 1.1e-245;
Matches 1471; Conservative 0; Mismatches 678; Indels 93; Gaps 2;

QY 80 TTAAGCAAGTGCAGGTGGCGCACCAAAACAGGAGTTGGTGGCCCAACATGCTCAACC 139
DB 117501 TCAACACACGCTGCGGCGCGCACCAACCAACCGTGACTGGTGGCCCAACGAGCTGAATC 117560

QY 140 TCGGCATCTTACGCCAACAATTCATCGCTATCGACCCCAACACACCGGATTTTGACTATG 199
DB 117561 TGAAGATTTTGACCAAGCATTTCTCGCTCTCCGACCCGATGGCGGAGGACTTTTGACTATG 117620

QY 200 CCGAAGAGTTTAAGAGCTAGACTGGCGACGGTTAAAGAGGACCTGGCGAGCGCTAATGA 259
DB 117621 CCAAGAGTTCAAGAGCTTGGACTTCGAAGCGGTCAGCAAGACCTTGCAGGAGTAATGA 117680

QY 260 CAGATTCAAGAGCTGGTGGCGACGAGATTACGGTCAATTTATGGCCCTCTTTTATACGCA 319
DB 117681 CCGGTCGACGAGCTGGTGGCGGCGGATTTTGGCCACTACGGCCGCTGTTTCAATTCGTA 117740

QY 320 TGGCGTGGCACAGCGCCGACCTACCGTATCGGTGATGGCGTGGTGGCGGTGGCTCGG 379
DB 117741 TGGCGTGGCACAGCGCGGTACTTACCGCACCGGTGATGGTCTGGCGGCGGCGCGG 117800

QY 380 GCTCAGAGCGCTTGGCGCTCTCAATAGCTGGCGACCAATGCCATCTGGATAAAGCAC 439
DB 117801 GTCAGAGCGGTTTGGCGCGCTCAATAGCTGGCTGACCAAGCTCAGCTTGGACAGGCGC 117860

QY 440 GCTTGGCTCTTTGGGCCATCAACAAAAATACGGTGGAAAAATCTCTGGCGGATCTAA 499
DB 117861 GTCGCTGATCTGGCGGCTCAAGCAGAGTACGGCGCGCAAGATTCTCTGGCGGACCTGA 117920

QY 500 TGATATCTACAGAACTAGCTCTGGAACATATGGCTTTAAACTTTTGGTTTGGAG 559
DB 117921 TCGTCTACCGGTAACTGGCGCTGGAGTGGCTATGGCTTCAAGACTTTCGGTTTTCGG 117980

QY 560 GTGGCAGACAGATGATGGAGCCTGGAAGAAGATGATATCTGGGAGCAGAAACCGAAT 619
DB 117981 GTGGTGGCGGATGCTGGAAACCGGAAGAGACGTGTACTGGGTTCCGAAACCCACT 118040

QY 620 GCGTGGGAGACAA----- 632
DB 118041 GCGTGGTGGCGAAGAGCGCTACGGCGCGCAGAAAGAAATGACGACGCGGCGAGCGCA 118100

QY 633 ----- 649
DB 118101 CCTGTGTGGCGGCTGAGAACCCAGCGCAAGAAAGACCGCCAGTGGCGGAAC 118160

QY 650 GAGAGCTCAAAATCCCTGGGAGCGGTACAAATGGGACTCATATGTAAACCCCGAAG 709
DB 118161 GTAACTGGAAACCCGCTGGCGCGCTGCAGATGGGCTTGATCTATGTGAACCCGGAAG 118220

QY 710 GACCCAAACGGCAGCCGACCTATCGCTGCTGCGCGTGATATTCGTGAGACTTTTGGCC 769
DB 118221 GCGCTGAAGCGGTGGCGGACCTGTGCTTTCGCGAGGAGTATCCGCGAAACCTTCGGCC 118280

QY 770 GAATGCCAATGAATGACGAAGAAACCGTGGCTCTCATACGGGTGGAGACACCTTCGGAA 829
DB 118281 GCATGGCCATGAATGACGAAGAGACCGTGGCGCTGATCGCGGTGGTACCGCTTCGGCA 118340

QY 830 AAACCCATGCTGCTCGGATCGGAGAAATATGTGGCGCGAGAGCTCGCGCGCAGGTA 889
DB 118340 TCTTGAATATCAACGTTCGGGCGAGAGCAAGCAGGTTGTGTTCACCGCAAGCCGGAACCC 119477

```

```

118341 AGACTCACGGTGCCTGGCGCGCGCGGCGGACAC- --GTGGGTCCAGAGCGGGAAGCAGCGGGTTC 118397
QY 890 TTCAAGAAATAGCCTGGGTGGAAAAACACCTAGCGCACCGGACACGCTGGGGATACCA 949
DB 118398 TTGAGCAACAGGCGTTTGGCTGGAGCAACAAATTTGGCACCGGTAAAGGTGGCGACCA 118457
QY 950 TCACAGTGGACTAGAGGCGCTGGACCAAGACCCCTACTCAATGAGGACAATAACTTTT 1009
DB 118458 TCACCAGTGGCTTGAAGTGAAGTGGACCTGACCTGACGCTACCAATGGAGCAACAGTACC 118517
QY 1010 TTGAAAACTCTTTTGGTTACGAGTGGAGCTTACCAAAAGTCCAGCTGAGCTTATCAGT 1069
DB 118518 TGAAGAACTCTTTTGGCTTGCATTTGGAACTGACCAAGAGCCGCGCGCGGCGCATCAGT 118577
QY 1070 GGAACCAAAAGACGCTGCCGGGCTGGCACCATACCGGATGCACATGATGCCAGCAAGT 1129
DB 118578 GGACGCTTAAAAACGGCGCTGGCGTGGCAAGATTCCCGGATGCCACGACCGGAGCAAC 118637
QY 1130 CCGACGCTCCATTTATGCTCACTACGGACCTGGCGCTGGCATGAGACCTGATTAGGAAA 1189
DB 118638 GGCATGCAACCGAGCATGCTGACCTCTGACCTGGCGTGGCTTTCGACCGCGCTACGAGC 118697
QY 1190 AAATTTCTCGACGGTACTATGAAACCCCTGATGAGTTTCAGATGCTTTTCGGGAAAGCAT 1249
DB 118698 AGATTTCTCGCGCTTCTTGGCCACCCCGACGCACTGGCCGCGTTCGCACGTGCTT 118757
QY 1250 GGTACAACTGACACAGAGATATGGGACCAAGAGTGGCTACCTTGGGACCAAGAGTGC 1309
DB 118758 GGTTCAGCTGACTACCGCGACATGGGCGCCACTGGCCCGCTACCTTCGCGCGGAAACAC 118817
QY 1310 CTCAGGAACCTCATCTGGCAAGACCTTATACCAGATGTAGCCATCTCTTTGTAGAG 1369
DB 118818 CGACCGAAGAGCTGTTTGGCGAGGACCCGATTCACAGACCTCATCTACGCTGGTGGATG 118877
QY 1370 AAAACGATATGAAAGGCTTAAAGCCAAATCTCTGGAATCGGACCTGACGGTAAGGAGC 1429
DB 118878 ATCAGATGTTGCCGCGCTCAAGGGCAAGATTCCTGACTCGGGCTTTTCGGTGTCAAC 118937
QY 1430 TGGTAAGCAGCGATGGGCTTCTGCTATCTACTTTTAAAGAACTGTACAAAGCGCGCGTG 1489
DB 118938 TGGTTTCAACCGCATGGCGGCGAGCTTCACCTTCCGCGCTCTGACAAAGCGCGCGGTG 118997
QY 1490 CCAACGGTGCAGTATACGACTGGCCCGCACAAAAGACTGGGAGTAAACACCTTCAGC 1549
DB 118998 CCAATGGCGGCGCTGCTGCTTGGCGCGCAGAAAGACTGGCGGCTCAACAGCTTCGAC 119057
QY 1550 AACTTGGCAGGCTACTCAAAACACTAGAAGTATCCAGGAGGACTTTAACACGAGCGCAAT 1609
DB 119058 AACTGGCAGCTGCTGAGTACGCTTGAAGTATTCAGAGCGAGTTCACGACGACTCAAT 119117
QY 1610 CAGATAACAAAGCAGTATCGTTGGCGGACCTGATGTGTGCGCGGCTGTGCGGGGTAG 1669
DB 119118 CGAACGGCAAGAAAGTGTCCATTTGCCGACCTGATGCTGTGCGCGGTGTGACAGCGCTG 119177
QY 1670 AAAAGCTGCAAAAGATGCTGGCCATGAGGTGACGCTGCTTTTCAACCGCGGACGAGCGG 1729
DB 119178 AGCAGCGCGAAGAAACGAGGTTCAGAGTTAACGGTTCCGTTTCACAGCGCGCGCGCGCG 119237
QY 1730 ATGGCAACCGCTGAGCAAAACCGGATGTGGAAGCTTTTCCGAAGCACTTAGAGCGCAGCGCTGACG 1789
DB 119238 ATGGCTCTCAAGAGCAGACAGATGTCGAGTCTGTTTCAAGCTTCCCTGAGCGGATGCGCGATG 119297
QY 1790 GCTTTAGAAACTACATTAACCGGAGCATAAAGTATCCGCTGAGGAAATGCTGCTAGAGC 1849
DB 119298 GCTTCGCAACTATCAGAAAGGCGGTTACAAGTATCGCGCGAGTCTGCTGCTGCTGAC 119357
QY 1850 GGGCGGAGCTTCTGTGCTTTTCGGCACCAAGAAATGACTGCTTTGGTGGGAGGATGCGGTG 1909
DB 119358 AGCGCAACTGCTGACTTTCGCTGCGGACAGATGACTGCTGCTGCTGCGGCTTCCGGG 119417
QY 1910 TACTGGCACCACTACGAGCTTCCGACATGAGTGTGTTACAAATAGCCGGGTGACG 1969
DB 119418 TCTTGAATATCAACGTTCGGGCGAGAGCAAGCAGGTTGTGTTCACCGCAAGCCGGAACCC 119477

```


Db 1143 GAGCGGAGGAGGCTTACTGGGTCCCGAGACCACCTGGCTGGACGACCGGGCTAC 1202
QY 640 GAAGGTGACGAGAGCTGGAATCCCTCGGAGCGCTACAAATGGAGCTCATCTATGTA 699
Db 1203 ACCGCGACCGGAGCTGGAGAACCGCTCGGCGGTCCAGATCGGCGCTGATCTACGTC 1262
QY 700 AACCCGAGGAGCAACGCGCAAGCCAGACCCCTATCGCTGCTGGCGTGATTCGTGAG 759
Db 1263 AACCCGAGGCGCGCAAGCCAGACCGGATCGGCGCGGCGCGGAGATCCGAGAG 1322
QY 760 ACTTTTGGCCGAATGGCAATGAATGACGAAGAAACCGTGGCTCTCATAGCGGGTGGAC 819
Db 1323 ACGTTCCGCCGATGGGATGAACGACGAGGACCGTTCGCTGATCGCGCGGCGCAC 1382
QY 820 ACTTTCGGAAGAACCCATGGTGTGCGGATCGGAGAAATATGTGGGGCGGAGAGCTGCC 879
Db 1383 ACTTTCGGAAGAACCCAGCGCGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1439
QY 880 GCCGAGGTATTGAAGAAATGAGCTGGGTGGAAACACCTACGGCACCGGACACGGT 939
Db 1440 GCCGCTCCCTCGAGGAGCGGGCTTGGGCTGGGCGAGCACTACGGCACCGGCAAGGCG 1499
QY 940 CGGATACCATCACAGTGGACTAGAAGCGCCTGGACCAAGACCCCTACTCAATGGAGC 999
Db 1500 CGGAGCCATCACTCCGCGCTGGAGTCACTTGGACCAAGCGCGGACCGAGTGGAGC 1559
QY 1000 AATAACTTTTGTAAACCTCTTTGGTTACGAGTGGGAGCTTACCAAAAGTCCAGCTGGA 1059
Db 1560 AACGGTCTTCAAGAACTCTTCGAGTACGAGTACGAGCTGGAGCAGAGCGCGCGCGC 1619
QY 1060 GCTTATCAGTGAACCAAGACGGTCCGGGCTGGCCACCATACCGGATGCAATGAT 1119
Db 1620 GCCACCAAGTGGTGGCGCAAGACCGCCCGAG-----ATCATCCCGACGCGCACGAC 1673
QY 1120 CCCAGCAAGTCGCGCTCCATTTATGCTCACTACGGACCTGGGCTCGCATGAGACCT 1179
Db 1674 CCGTGAAGAGACCGCGCGGNTGCTACACCGGACCTGCTCCGCGCTTGACCCG 1733
QY 1180 GATTACGAAAAATTTCTGACGCTACTPATGAAACCTCGATGATTTGCAAGTCTTC 1239
Db 1734 ATCTACGAGCGATCTCCGCGGTTCTACGAGAACCGGAGGAGTTCGCGGACCGCTTC 1793
QY 1240 CGAAAGCATGTAACAACTGACACAGAGATATGGACCAAGGTGCGGTACTCTGGA 1299
Db 1794 GCCCGCGCTGTACAACTCACCACCGTGCATATGGCGCGGAAGTCTGTACTCTCGC 1853
QY 1300 CCAGAGTGCCTCAGAGAGACCTCATCTGCGAAGACCTTATACAGATGTAAAGCATCCT 1359
Db 1854 CCGGAGTCCCGAGGAGACGCTGCTGTGGCAGGACCGCTCGCGGCGGAGGCGAG 1913
QY 1360 TTGTAGAGAAACGATATTGAAGGCTTAAAGCCAAATCTCTGGAATCGGAGTGAAG 1419
Db 1914 CTGATCGACGACGCGGACATGCCATCTCAAGACCAAGCTGCTGGAGTGGCGGCTGTC 1973
QY 1420 GTAACGAGCTGTGAAGCAGCATGGCTTCTGCATCTACTTTTGAAGTCTGACAG 1479
Db 1974 GTGTCCCAAGTGTGTCACCACTGCGTGGCGGTCCGCTCCACTTCGCGGCGGACAG 2033
QY 1480 CGCGCGGTGCAACGCTGACGATACGATGCGCCGCGGCGGCGGCGGCGGCGGCGG 1539
Db 2034 CGCGCGGTGCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2093
QY 1540 AACCTCAGCAACTTGCAGGCTACTCAAAACACTAGAGGTATCCAGGAGGACTTTAAC 1599
Db 2094 GACCCCGACAGCTGGCGAGGTGTGCGGACCCCTGGAGAGCGTCCAGCAGGAGTTCAAC 2153
QY 1600 CAGGCGCAATCAGATAACAAGCAGTATCTGTGGCGGACCTGATTTGCTGCCCGGTGT 1659
Db 2154 ---GCTCTCTCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2210
QY 1660 GCGGTGTAGAAAAAGCTGCAAAAGATGCTGGCCATGAGGTGCGAGTGCCTTCAACCG 1719
Db 2211 GCCGCGGTGGAGAGCGCGCCNAGGAGCGGCTTCGAGATCCAGGTGCCCTTACCCCG 2270

QY 1720 GGACGAGCGGATGCCACCGCTGAGCAAAACCGATGTGAAGCTTTTGAAGCACTAGACCA 1779
Db 2271 GCGCGTGTGACGCGGACCGAGGAGCAGACACCGAGCTCGAGTCTTCGAGGCGCTGGAGCG 2330
QY 1780 GCGCTGACGCGCTTAGAACTACATTAACCGGAGCATAAAGTATCGCTGAGGAATG 1839
Db 2331 ACCGCGGACGCGTTCGCAACTACTCGGCAAGGCAACCGGCTGCGGCGGCGGAGTACCTG 2390
QY 1840 CTCGTAGACCGGCGGCGAGCTTCTGCTGCTTTTCGCGCACGAGAAATGACTGCTTTGTAGGC 1899
Db 2391 CTGCTCGACAAGCGCACTGCTGAACCTGAGCGCGCCCGAGATGACGGTGTGCTGCTGCTG 2450
QY 1900 GGTATGCGTGTACTTGGGCGCAACTACGACCGGTTGCGAGCATGAGTGTTCACAAATG 1959
Db 2451 GGTGTGCGGCTCTCGGCGCAACCGACGAGCTGCGAGCTCGGCGTCTTCCACCAAGACG 2510
QY 1960 CCGGCTGAGCTATCCAATGACTTCTTGTAAACCTGTAGACCTCAACACTAAATGGCGA 2019
Db 2511 CCCGCGTCTCACCAAGACTTCTTGTCAACCTGTCTGACATGGGAGCAGCGTGGAG 2570
QY 2020 GCCAGCATGAATCAGACAAAGTTTGAAGCGCAGAGACTTCAAAACTGGCGAACTAAAG 2079
Db 2571 GCGACCTCCGAGGACCAAGCAGCTTCGAGGCGCGGAGCGCGGCGGAGGTGAAG 2630
QY 2080 TGGAGTGGCACCGGCTGAGACTGATCTTCGGATCCAAATTCGAGCTTAAGAGCCCTCGCA 2139
Db 2631 TGGCGCGGCGGAGCGGAGCTGCTTCGCTTCCAACTCGAGCTCGCGCGCTCGCC 2690
QY 2140 GAAGTGTACGCTGTGCAGATTCGAGAAAAGTTTGTAAAGTATTTGTGAGGCCCTGG 2199
Db 2691 GAGGTCTACGCGCAGCAGCGCCAAAGGAGAGTTCGCAAGGACTTCTGCGCGCGTGG 2750
QY 2200 GCCAAAGTAATGGACCTGGACCGGTTGATCT 2231
Db 2751 CACAAGTCTAGGAGCGGACCGGTTGACCT 2782

RESULT 9
LOCUS AY040244
DEFINITION Burkholderia pseudomallei ATP-dependent DNA helicase protein RecG (recG) gene, partial cds; and oxidative stress regulatory protein oxyr (oxyr), catalase-peroxidase protein KatG (katG), and DpsA (dpsA) genes, complete cds.
ACCESSION AY040244
VERSION AY040244.3 GI:22711992
KEYWORDS Burkholderia pseudomallei
SOURCE Burkholderia pseudomallei
ORGANISM Burkholderia pseudomallei
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae; Burkholderia; pseudomallei group.
REFERENCE 1 (bases 1 to 4910)
AUTHORS Loprasert,S., Sallabhan,R., Whangsuk,W. and Mongkolsuk,S.
TITLE The Burkholderia pseudomallei oxyr gene: expression analysis and mutant characterization
JOURNAL Gene 296 (1-2), 161-169 (2002)
REFERENCE 2 (bases 1 to 4910)
AUTHORS Loprasert,S., Sallabhan,R., Whangsuk,W. and Mongkolsuk,S.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-2001) Lab. of Biotechnology, Chulabhorn Research Institute, Vipavadee, Bangkok 10210, Thailand
REFERENCE 3 (bases 1 to 4910)
AUTHORS Loprasert,S., Sallabhan,R., Whangsuk,W. and Mongkolsuk,S.
TITLE Direct Submission
JOURNAL Submitted (07-JUN-2002) Lab. of Biotechnology, Chulabhorn Research Institute, Vipavadee, Bangkok 10210, Thailand
REMARK sequence update by submitter
REFERENCE 4 (bases 1 to 4910)
AUTHORS Loprasert,S., Sallabhan,R., Whangsuk,W. and Mongkolsuk,S.
TITLE Direct Submission
JOURNAL Submitted (04-SEP-2002) Lab. of Biotechnology, Chulabhorn Research Institute, Vipavadee, Bangkok 10210, Thailand

```
REMARK      Sequence update by submitter
COMMENT      On Sep 4, 2002 this sequence version replaced gi:22657796.
FEATURES
source       1. .4910
             /organism="Burkholderia pseudomallei"
             /mol_type="genomic DNA"
             /db_xref="taxon:28450"
             <1. .183
             /gene="recG"
             <1. .183
             /gene="recG"
             /note="similar to Escherichia coli ATP-dependent DNA
             helicase RecG c-terminal region"
             /codon_start=1
             /transl_table=11
             /product="ATP-dependent DNA helicase protein RecG"
             /protein_id="AAK72464.1"
             /db_xref="GI:16580119"
             /translation="EFLGARQSGAAMLRPADLENDGWLIEPARDAAARLIAAHPEVVA
             QHARWIGAREQYLKA"
             319. .1278
             /gene="oxyr"
             319. .1278
             /gene="oxyr"
             /note="global regulator which belongs to the Lysr family
             of transcription factors"
             /codon_start=1
             /transl_table=11
             /product="oxidative stress regulatory protein Oxyr"
             /protein_id="AAK72465.1"
             /db_xref="GI:16580120"
             /translation="MTLTELKVIYAVAREHFRGAAEAFVSOPITLSVAIKKLEDELN
             VOIFERGASEVSTVPQDQIVTQARVLEQTFAIKRAKQGDPLIGPRLGLVIYIG
             PVLPLTFLVMRRVPPMLQENYTLKLELLKQGEIDAAIMALPPFETGLMVRAL
             YDEPFVVALPAGHPWEKRRKIDAADLKQETMLLGNHCFRDRHVLGVCPELMHFSOTA
             DGIOKTFEGSLETIRHVASGVGTLVLRMSVAEVPAGSGPADLLSYVPEDPVP
             DRVVLWRKSTRMPAIDAISEAIAACGLPGVTKLIDIPATN"
             1704. .3950
             /gene="katG"
             1704. .3950
             /gene="katG"
             /note="similar to Escherichia coli catalase-peroxidase
             KatG"
             /codon_start=1
             /transl_table=11
             /product="catalase-peroxidase protein KatG"
             /protein_id="AAK72466.3"
             /db_xref="GI:22711193"
             /translation="MPGSDAGPRRRGVHEQRNRNMSNEAKCPHQAGNGTNRDWWP
             NQIDUSILRRKSLSDPMGKDNFYAQAFKELDLAAVKRDLHALMTTSQDWWPADFGHY
             GGLFTMAHWSAGTYRTADRGAGGEGQRFAPLNSWPDNANLDKARLLIPIKQYK
             RAISWADLLILGNVALESMPKFTFGAGGRADTWEPDYWGSEKIMLELSGGPNRS
             YSGDROLENPLAAVQMLIYVNPPEGDPGPNPVAARDIRDTFARMAMNDEETVALIA
             GGHTFGKTHGACGASNVGLAEPEAAGTEAOLGLWKSAYRTGKGADATTSGLEVTWTTTP
             TORSHNEFFENLGYEWELTKSPAGAHQWAKGADAVIPADFPSPKRRHTMLTDLSL
             REDPAYEKSRAPHENPEGFADAFARAFKLRHDMGPRARYLGPPEVPAEVLMDQPI
             PAVDHLIDAAAEELKARVLSGLTVSVTAWAAASTFPGSKRGGAANGARTLA
             PQKDEANQPEGLAAVLETLARTAFNGAQGGQKQVSLADILVLACGAGVQBAARNA
             GHAVTPVAPGRADASQEQTDVESNAVLEPVADGPNYLRKGYRVPAYVLLVDKAQLL
             TILSAPMTVLLGRLVLANVQSRHGVTEAREOALTNDFFVNLIDMGTEKPTADA
             DYFEGEDRATGELKWTGTRVDLVFGSHSOLRALAEVYGSADAQENFVRDFVAVNKNVM
             NLDKFDLA"
             4422. .>4910
             /gene="dpsA"
             4422. .4910
             /gene="dpsA"
             /note="non-specific DNA binding protein"
             /codon_start=1
             /transl_table=11
             /product="DpsA"
             /protein_id="AAK04250.1"
             /db_xref="GI:22657798"
             /translation="MAKKSNAQINIGISDKDRKKIAGLSRLADTYLLYLKTHNFH
```

```
WNVTPGFNTLHLMFEQYNELWLDVAERITLGVVAPGTYYREFAKLSSIPEDG
VPAAEEMIRQLVEGHEAVVTRARAIFFDADAASDEPTADLLIQLRQTHEKTAWMLRSL
LA"
BASE COUNT      834 a 1590 c 1736 g 750 t
ORIGIN
Query Match      41.6%; Score 932; DB 1; Length 4910;
Best Local Similarity 65.8%; Pred. No. 1.5e-240;
Matches 1428; Conservative 0; Mismatches 715; Indels 27; Gaps 4;
QY 77 CGCTTAAAGCAAGTCAGGTGGCGGACCAACAAAACAGGATTCGTGCCCAACATGCTCA 136
Db 1786 CGTTCCATCAAGCCGAGGCAACGACGCGTCAACCGGACTGTCGCCCAACATGCTCA 1845
QY 137 ACCTCGGCATCTTACGCCACATTCATCGGTATCGGACCCAAACAGACCCGGATTTGACT 196
Db 1846 ACCTGAGCATCTTCGATCGGCACTCGTCGTCTGCCATCGGATGGGCAAGGATTTCAACT 1905
QY 197 ATGCCGAAGAAGTTTAAAGAAGCTAGATCTGGCAGCGGTATAAAAAGACCTTGGCAGCGCTAA 256
Db 1906 ACGCGCAGCGGCTTCGAGAGCTCGACCTCGCGCGGCTGAAGCGCGACCTCCACGCGCTGA 1965
QY 257 TGACAGATTCACAGGACTGTGGCCAGCAGATTAACGGTCATTAATGCGCCCTTTCTTTATAC 316
Db 1966 TGACGACGTGCGCAGGACTGTGGCGCGGCGGATTCGCGCCACTACGCGCGGCTTTCATCC 2025
QY 317 GCATGGCGTGGCACAGCGCGGCACCTACCGTATCGGTATGCGGTGGTGGCGGTGGCT 376
Db 2026 GCATGCAATGGCACAGCGCGGCGACGTACCGCAGCGCCGACGCGCGCGCGCGCGGCG 2085
QY 377 CCGGCTCACAGCGCTTCGGCGCTCTCAATAGCTGGCCAGCAATATGCAATCTGGATAAAG 436
Db 2086 AAGGCGCAGCAGCTTCGGCGCTCTCAACAGCTGGCGCCGACAAACGCGAACTCGACAAGG 2145
QY 437 CAGCGTGTCTTCTTGGCCCATCAAAACAAATACGGTCGAAAAAATCTCCTGGCGCGATC 496
Db 2146 CCGCGCGGCTGTCTGTGGCGGATCAAGCAGAAAGTACGCGCGCGGCATCTCGTGGCGGACC 2205
QY 497 TAATGATATCTACAGGAAACGTAAGCTCTGGAACATATGGCTTTTAAACATTTTGGTTTG 556
Db 2206 TGCTGATCTGAGGCGCAACGTCGCGCTCGAATCGATGGCTTCAAGACCTTCGGCTTCG 2265
QY 557 CAGTGGCAGACAGATGTATGGGAGCGCTGAAGAAGATGTATATCTGGGAGCAGAAACCG 616
Db 2266 CCGCGCGCGCGCGCAGCAGCTGGGAGCC---CGAGGACGCTCTACTGGGCTCGGAAAGA 2322
QY 617 AATGGCTGGGA-----GACAAGCGCTTGAAGTGAACCGAGACTCGAAA 661
Db 2323 TCTGGCTGAACTGACGCGGCGGCCGACAGACGCGCTATTTCGGCGCAGCCGCTCGAGA 2382
QY 662 ATCCCTGGGAGCGGTACAAATGGGACTCATCTATGTAACCCGAGGACCCCAACGCGCA 721
Db 2383 ACCCGCTCGCGCGGTGAGATGGGCTCATCTACCTGAATCGGAAAGCCCGGACGCGCA 2442
QY 722 AGCCAGACCTATCGCTGTCTGCGGTGATTCGTGAGACTTTTGGCCGAATGGCAATGA 781
Db 2443 ATCCCGATCGGTGCGCGCGCGCGGACATTCGTGACACCTTTCGCGCGCATGGCGATGA 2502
QY 782 ATGACCAAGAAACCGTGGCTCTCATATACGGGTGGACACACCTTTCGAAAAAACCCATGGTG 841
Db 2503 ACAGCAAGACGCGTTCGCGCTGATCGCGGGGGGCCACACGTTTCGCAAGACGACGCGG 2562
QY 842 CTGCCGATCGGAGAAATATGTGGGCGGAGAGCTCGCCCGCGCAGGATTTGAAGAAATGA 901
Db 2563 CGGGGCGCGGTCGA---ACGTGGCGCGGAGCGCGGCGCGGCGCATCGAAGCCGAGG 2619
QY 902 GCTTGGGTGGAAAAACACCTACGGCACCGGACACGGTGGGATACCATCACAGTGGAC 961
Db 2620 GCCTCGGCTTGAAGAGCGGCTACCGCACGGGCAAGGCGCGGACGCGGATCACAGCGGCG 2679
QY 962 TAGAAGCGCGCTCGACCAAGACCCCTTACTCAATGGAGCAATTAATTTTGTAAAAACCTCT 1021
Db 2680 TCGAAGTCACTGAGACGACGACCGCGCAGTGGAGCCACAACTTCTTCGAGAACCTGT 2739
```


QY 634 CGCTATGAGGTGACCGAGAGCTCGAAATCCCTGGGAGCGGTACAAATGGGACTATC 693
Db 838 CCCTACGCGGAGCGGTGAGTGGACTCGCCCCCTGGGCGGGTCCAGATGGGCGCTGATC 897
QY 694 TATGTAAACCCCGAAGGACCAACGCGCAAGCAGACCTATCGCTGCTGCGCGTGATATT 753
Db 898 TACGTCAATCCGAGGCGCGAAGCAACCCGACCCGCTGGCTTCGGCGCGGACAT 957
QY 754 CGTGAGACTTTTGGCCGAATGCAATGATGACGAAGAAACCGTGCGTCTCATACGCGGT 813
Db 958 CGCGAGACTTCGCCCGGATGCCATGAATGACGAGGAGACGCTGCGCTGATCCGCGC 1017
QY 814 GGACACACTTCGGAANAACCATGCTGCTCGGATGCGGAGAAATATGTGGCGCGAG 873
Db 1018 GCGCACACTTCGGAAGGCTCACGGCGGCGGACGC---CTCCCTGTGGGCGTTGAG 1074
QY 874 CCGTCCGCGCGCAGGTATTGAAGAAATGAGCCTTGGGGTGGAAAAACACCTACGGCACCGGA 933
Db 1075 CGGAAGCGCGCGCATCGAGGCCCAGGGCTTTGGGTGGCGGACGAACGACGCGACGGC 1134
QY 934 CACGCTGGGGATPACCATCACCACTGAGTACGAGCGGCTGGACCAAGACCCCTACTCAA 993
Db 1135 AAGGGCCCGCGCCCATACCGCGCGCGCGGAAGTGATCTGGACCCAGACGCGCCACGCGC 1194
QY 994 TGGAGCAATACTTTTGAANAACCTCTTTGGTTACGAGTGGAGCTTACCAAAAGTCCA 1053
Db 1195 TGGAGCAACCACTTTTCGACAACTGTTCAGTACGAGTGGGAGCTGAACCAAGCCCC 1254
QY 1054 GCTGGAGCTTATCAGTGGAAACCAAAAGACGCTGCGGGCTGGCACCATACCGGATGCA 1113
Db 1255 GCGCGCGCAACGATGCGAGCGCAAGAA-----CGCCCGCGCGACATTCGAGCGCC 1308
QY 1114 CATGATCCGAGAGTCCGACGCTCCATTTATGCTACTACGAGACCTGGCGCTGGCGATG 1173
Db 1309 TTCGATCCGACAAAGACGACGCTTCGCGGATGCTCAGCTCGGACCTCGCCCTGGCTTT 1368
QY 1174 GACCTGATTAGCAAAATTTCTCGAGGTTACTATGAAACCCCTGATGAGTTTCAGAT 1233
Db 1369 GACCCGGCTACGAGAAGATCTCGCGCGGTTCTACGAGAACCCGAGCAGTCCGCCAC 1428
QY 1234 GCTTTCCGGAAGCATGTGTACAACTGCACACAGAGATATGGGACCAAGTGGCGTAC 1293
Db 1429 GCCTTCGCGCGCTGTTCAAGCTGACCCACCGGACATGGCCCGATCGCGCGCTAT 1488
QY 1294 CTGGGACCAAGATGCTCCGAGAACCTTCTGGCAAGACCTATACCAAGATGATAGC 1353
Db 1489 CTGGGCGCGCTGCTCCGAAAGGAGGAGTGTCTGGCAGGACCCGATCCCGCCCTTGAC 1548
QY 1354 CATCCTCTGTAGACAAACCATATTGAAGGCTTAAAGCCCAAAATCCTTGAATCGGA 1413
Db 1549 CACCCGCTGGCGACGACAGGACATCGCGGCGCTCGAAGGCCAAGATCCTGGCCACCGC 1608
QY 1414 CTGACGGTAAGCGAGCTGTAAAGCAGCGGATGGGCTTCTGCACTACTATTTAGAACTCT 1473
Db 1609 CTGTGCGCTCGGACCTCGTCTCGACCGCTCGGCTCGGCTCGACCTATCGCCAGTGG 1668
QY 1474 GACAAGCGCGGCTGCCAACGGTGCACGTATACGACTGGGCCCCCAAAAGACTGGGAA 1533
Db 1669 GACAAGCGCGGCGGCGCAATCGCGCGGATCCGCTGCGCCCGCCAGAAAGACTGGGCG 1728
QY 1534 GTAAACACCCCTCAGCAACTTCCAGGGTACTCAAAACACTAGAGGTATCCAGAGGAC 1593
Db 1729 GTCAACAACCCCGGCTCTGCGCAAGGTGCTGGGCGGCTCGAAGGGCTCCAGAGGAC 1788
QY 1594 TTTTAACCGCGCAATACAGATAACAAAGCAGTATGTTGGCGGACCTGATTTGCTGGCG 1653
Db 1789 TTTCAAGCTTCGCGCGGTGGCGCAAGAGATCTCGCTGGCGGACTTGATCTGCTGGCG 1848
QY 1654 GCGTGTGGGGTGTAGAAAAGCTGCAAAAGATGCTGGCCATGAGGTGACGCTGCTTTTC 1713
Db 1849 GCGCGGCGCGGCATCGAAAAGCGCGCAAGGACGCGGAAACGAGCTACGCTGCGCTTC 1908
QY 1714 ACCCGGGACGAGCGGATGCCACCGCTGAGCAAAACCGATGTGGAAGCTTTCGAAGCACTA 1773

Db 1909 GCGCGGCGCGCATGGACGCTTCGGCCGAACAGACCGAGCTCACTCTTCGAGGCCCTG 1968
QY 1774 GAGCAGCGCGCTGACGGCTTTTGAAGAACTACATAAACCGG---AGCAATAAGTATCCGCT 1830
Db 1969 GAGCCCGCTCAGACGGCTTCGCAACTATCGCGGCGCGGCAAGCACTACATGCGCCGG 2028
QY 1831 GAGGAATGCTGTAGACCGGCGGAGCTTCTGTGCTTTTCGGCAACCAAGAACTGCT 1890
Db 2029 GAAGAGCGCGCTGTGAGACCGCGCCAGCTGTGGGCTGTGCGGTCGGAACGAGGCTG 2088
QY 1891 TTGTAGCGGTATGCTGTACTGGGCAACCACTACGAGCGTTCGACGATGGAGTGT 1950
Db 2089 CTGGTGGCGGCTCGCGTGTGGCGCAATTCGAGCGCTCCAGGACGGGTGTC 2148
QY 1951 ACAATAAGCGCGGTGAGCTATCAATGACTTCTTTGTAAACCTCTAGACCTCAACT 2010
Db 2149 ACCAACCGCGCGCGCTGAGCAATGACTTCTTCGTAAGCTCTCAGCATGGAAC 2208
QY 2011 AATGGCGAGCGAGCATGATGATGACAAAGTTTTTTGAAGGAGAGATTCNAAACTGGC 2070
Db 2209 ACCTGGACCGCGACGG-----CGGCCAACGCTTCGCGCGCATGACCGCAAGTCCAGC 2262
QY 2071 GAAGTAAAGTGGAGTGGCACCGCGGTAGACCTGATCTTCGGATCCAATTCGAGCTAGA 2130
Db 2263 GAGCGCGTTGGACGCAACCGCGCTGACCTGATCTTCGGCTCCGACGCGAAGTGGG 2322
QY 2131 GCGCTTCGAGAGTGTACGGCTGTGAGATTTCTGAAGAAAAGTTTGTAAAGATTTTGT 2190
Db 2323 GCTTCGCGGAGTGTATGCTGCGCGGACTCGCAGGAGAGTTCGCTGCGACTCGTC 2382
QY 2191 AAGGCTGGGCAAGTAAATGAGCTGGACCGGTTGATCTG 2232
Db 2383 ACCGCTGGAACAGGTGATGAACCGCACCGACCGCTGGATCTG 2424

RESULT 11
RLE438039

LOCUS Rhizobium leguminosarum bv. phaseoli panB (partial), oxyR and katG
DEFINITION genes.

ACCESSION AJ438039

VERSION AJ438039.1 GI:19338426

KEYWORDS catalase-peroxidase; katG gene; ketopantoate hydroxymethyltransferase; oxyR gene; panB gene.

SOURCE Rhizobium leguminosarum bv. phaseoli

ORGANISM Rhizobium leguminosarum bv. phaseoli

Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.

REFERENCE 1

AUTHORS Chen, J.Y.

TITLE Molecular Genetics and Physiology of a Catalase-Peroxidase of

Rhizobium leguminosarum bv. phaseoli

JOURNAL Thesis (1999) Department of Biology, Imperial College, University

of London, London, United Kingdom

REFERENCE 2 (bases 1 to 4754)

AUTHORS Williams, H.D.

TITLE Direct Submission

JOURNAL Submitted (01-MAR-2002) Williams H.D., Biological Sciences,

Imperial College, Imperial College Road, London, SW7 2AZ, UNITED

KINGDOM

FEATURES

source

1. .4754
/organism="Rhizobium leguminosarum bv. phaseoli"
/mol_type="genomic DNA"
/strain="4292"
/db_xref="taxon:385"
/country="United Kingdom"
/note="biovar: phaseoli"
3. .812
/gene="panB"
<3. .812

	1416	GACGGTAAGCGAGCTGGTAAAGCAGCCGATGGGCTTTGTGCATCTACTACTTTTTAGAAACTCTGA	1479
QY			
Db	3259	TACCCTGTCAGGAACCTGGTCTGCAACGCCCTGGCGCTCGACCTTCGCGGGCTCCGA	3318
QY	1476	CAGCGCGCGGTGCCAACCGGTGCAGGTATACGACTGGCCCCACAAAAGACTGGGAAGT	1535
Db	3319	CAGCGCGCGCGCCCAATTGGCGCGCGCTCGCTTGCCCGCAGAAGNCTGGGACGC	3378
QY	1536	AACAACCCCTCAGCAACTTGCACAGGTACTCAAACACATAGAAGTATCCAGGAGGACTT	1595
Db	3379	CAATCAGCCGCCAGCTGCCAAGGTGCTCGGCGTTCTCGAAGCNTCCAAAAGACTT	3438
QY	1596	TACCAAGGCGCAATCAGATAACAAGACGATGCTGTGGCCGACCTGATTTGCTGTCGCCCG	1655
Db	3439	CAACGCGCCGACAGCCGGGGCCAAAGAAGATCTCGCTGGCCGACCTCATGCTTCTGCCCG	3498
QY	1656	CTGTGCGGTGTAGAAAAAGCTGCAAAAGCATGCTGCCATGAGGTGTCAGTGCTTCCA	1715
Db	3499	TGCTGCCGTGTGAGAAAGCGGCGGCGGCAACCGCTCAGCGTGCCTTTCAC	3558
QY	1716	COGGGACGAGCGGATGCCACCGCTGAGCAAAACCGATGTGGAGCTTTCCGAAGCATAGA	1775
Db	3559	GCGGCGCCGATGATGCTTCCGAAGCCAGACGCGCATTCCTTCCGACGCTCGA	3618
QY	1776	GCCAGGCGCTGACGGCTTTAGAAACTACATTTAAACCGGAGCATAAAGTATCGB-----C	1829
Db	3619	GCCGCGTATCGACGGCTCCGCAATTTACTTGAACGGCAAGCGCATCAGTTTATGAAGC	3678
QY	1830	TGAGGAAATGCTGTAGACCGGGCGCAGCTTCTGTGCGCTTTCGGCACCAAGAAATGACTGC	1889
Db	3679	GGNAGAAGCGCTGCTGACCGCGCCGCTTGTGTGACGCTTGACCGGACCTGAATGACCGC	3738
QY	1890	TTTGTGAGCGGTATGCTGTACTGCGGCACCAACTACGACGGTTCGACGATCGGAGTGT	1949
Db	3739	CCTGCTGCGCGCTCGCGTGTG-----AAGCGCGCGCCCGGACGCGGCTCTT	3792
QY	1950	TACAAATAGCCGGTTCAGCTATCCAAATGACTTCTTTGTAACCTGCTAGACCTCAACAC	2009
Db	3793	CACCTCGCTCCGAGACCTGACGAACGACTTCTTCGTCGAACCTGCTCGCATGGCCAC	3852
QY	2010	TAAATGGCAGCGAGCGATGAATCAGACAAAGATTTTGAAGGCAGAGACTTCAAAACTGG	2069
Db	3853	CGAGTGGTTCCTGTGCGCGGCAAGGAGCTTCTATGAAGCCGCGACCGCAAGACCGG	3912
QY	2070	CGAAGTAAAGTAGAGTGGACCGGGTAGACCTGATCTTCGATCCAAATCCGAGCTAAG	2129
Db	3913	CGCTGCAAGTAGACCGGACCGCGCTTGACCTGATCTTCGGCTCGCACTCGCAGCTCG	3972
QY	2130	AGCCCTCGCAGAGTGTACGGCTGTGCAGATCTGAAGAAAAGTTTGTAAAGATTTCT	2189
Db	3973	CGCTTCGCCGAAGTCTACGCCGAGCCGACCGCAAGGGGAGTTCGTCAATGACTTCGT	4032
QY	2190	GAAAGCCTGGCCCAAGTAAATGGACCTGGACCGGTTTGATCT	2231
Db	4033	CGCCGCTTGGAAACAAGTCTATGAACCGGACCGCTTCGACCT	4074
RESULT	12		
LOCUS	SC0939106/c		
DEFINITION	Streptomyces coelicolor A3(2) complete genome; segment 3/29.		
ACCESSION	AL939106 AL096837 AL121746 AL132662 AL132856 AL132973 AL132991 AL133171 AL133424 AL136149 AL136502 AL137165 AL445945 AL449216 AL591823 AL593822 AL645882 AL939106.1 GI:24413728		
VERSION			
KEYWORDS			
SOURCE	Streptomyces coelicolor A3(2)		
ORGANISM	Streptomyces coelicolor A3(2)		
BACTERIA:	Actinobacteria; Actinobacteridae; Actinomycetales;		
STREPTOMYCINEAE:	Streptomycetaceae; Streptomyces.		
REFERENCE	1		
AUTHORS	Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, P.M., Challis, G.L.,		

TITLE	Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)
JOURNAL	Nature 417 (6885), 141-147 (2002)
MEDLINE	1296410
PUBMED	12000953
REFERENCE	2 (bases 1 to 314100)
AUTHORS	Bentley,S.D.
TITLE	Direct Submission
JOURNAL	Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk
COMMENT	On or before Oct 26, 2002 this sequence version replaced gi:20520920, gi:20520898, gi:20520901, gi:20520903, gi:20520904, gi:20520905, gi:20520906, gi:20520907, gi:20520926, gi:20520909, gi:20520910, gi:20520805, gi:20520706, gi:20520789, gi:20520843.
FEATURES	Location/Qualifiers
source	1..314100 /organism="Streptomyces coelicolor A3(2)" /mol_type="genomic DNA" /strain="A3(2)" /db_xref="taxon:100226" 53..56 67..945 /gene="SCO0550" /note="synonym: SCF11.30" 67..945 /gene="SCO0550" /note="SCF11.30, possible membrane protein, len: 292 aa; similar to TR:09ZBQ4 [EMBL:AL034447] Streptomyces coelicolor putative transmembrane protein, 396 aa; fasta scores: opt: 677 z-score: 505.6 E(): 8.6e-21; 43.7% identity in 309 aa overlap. Contains possible hydrophobic membrane spanning region" /codon_start=1 /transl_table=11 /product="putative membrane protein" /protein_id="CAB59607.1" /db_xref="GI:6137052" /translation="MGSRRAAPRSRSHAPRGCLRSGAALGALGATLTCAGVGTITLAV GKDDGSDRVTANADQAPNVEAANGGAAPLASASPTAGTSRSPSPSAEPKRTPT TKTASPTAPKASAPKAPVRAATKSGGDRSGSGSGSGSGSGSGSGSGSGSGSGS GSGTGGAGAAQYALVNEERAAAGCSFVTANRLTRADDYSDVMASSGVMHITPGDG STMASRVEAAGYQWSALQIAGQQAASVMEAWMNSPCHRENILNCSFKELGVGVGH FDGGGPWTQDFGTGR" complement(1031..2359) /gene="SCO0551" /note="synonym: SCF11.31c" complement(1031..2359) /gene="SCO0551" /note="SCF11.31c, possible histidine kinase protein, len: 442 aa; similar to TR:050497 [EMBL:AL020958] Streptomyces coelicolor two-component system sensor, 481 aa; fasta scores: opt: 401 z-score: 451.3 E(): 9e-18; 30.2% identity in 367 aa overlap. Contains Pfam matches to entries PF00512 signal, histidine kinase and PF00672 DUF5, domain found in bacterial signal proteins. Contains possible N-terminal signal peptide sequence" /codon_start=1 /transl_table=11 /product="putative histidine kinase protein" /protein_id="CAB59608.1" /db_xref="GI:6137053" /translation="MRRALAGIALAVTSMVALSFLVPLALLVREQARDRVTTAAEQRA AALSPLWLTTRADYQQVAEIGSDAQIVRLPDGGFTVGTTHAPAEADLVRGVRGT LAVDTYADYATQPVVLRDRVAVVPAADLTGVTASGVMALLAVGLVGSVL

VADRLGARVVRSSRLKRALGSLGDLGVRRVDPPELQEAAGAENFWADRVVDLL
 AVRELVDLSHGRUKTLEALYLEADRMGATPPSALKRYTEAGQLERELDSLIAAATPD
 LAAQLAGTGAAGTACPDVAEAVMRDLDFWLSAAQQDRPFERSLTFRPAFVAPEDP
 LAAVVDALIGNVFHTPGQTRFVVRVRRDRHVRLLTVDDAGPGVADPEAALTGCVSG
 GSTGLGDIIVARAATDAGELTITRAPLGGARVTVSPALADARAAGDGR"
 complement(1070..1672)
 /gene="SC00551"
 /note="Pfam match to entry PF00512 signal, Histidine
 kinase, score 39.60, E-value 7.1e-08"
 complement(1712..1921)
 /gene="SC00551"
 /note="Pfam match to entry PF00672 DUF5, Domain found in
 bacterial signal proteins, score 36.90, E-value 4.5e-07"
 complement(2356..3072)
 /gene="SC00552"
 /note="synonym: SCF11.32c"
 complement(2356..3072)
 /gene="SC00552"
 /note="SCF11.32c, possible response regulator, len: 238
 aa; similar to TR:CAB54568 (EMBL:AJ006392) Streptococcus
 pneumoniae response regulator Rro2, 234 aa; fasta scores:
 opt: 583 z-score: 693.9 E(): 2.8e-31; 43.2% identity in
 227 aa overlap. Contains Pfam matches to entries PF00486
 trans_reg.C, transcriptional regulatory protein and
 PF00072 response_reg, response regulator receiver domain"
 /codon_start=1
 /transl_table=11
 /product="putative response regulator"
 /protein_id="CAB59609.1"
 /db_xref="GI:6137054"
 /translation="MTDRLLCLPMASVLVDEDDPVIIRAALIEVLTHGYAVKTAHQGF
 EALDITQSPDIDVLGLDGLDLRMIRGISRPVLVATARDDEFIIRLLNA
 GADYVMKPFSSGQLAAALVRRSPAGLRAARPPVLRVLDIPTARTAHLAGR
 ELPTREFFDLLAYLAHAHQVMSRQRIILAEVWQPYVEDQTVDVHLSALRRKMGKA
 RKPRYLTVRGIGIKLVSAIP"
 complement(2380..2595)
 /gene="SC00552"
 /note="Pfam match to entry PF00486 trans_reg.C,
 transcriptional regulatory protein, C terminal, score
 93.00, E-value 1.7e-26"
 complement(2710..3042)
 /gene="SC00552"
 /note="Pfam match to entry PF00072 response_reg, Response
 regulator receiver domain, score 123.00, E-value 5.6e-33"
 complement(3090..3093)
 complement(3099..3536)
 /gene="SC00553"
 /note="synonym: SCF11.33c"
 complement(3099..3536)
 /gene="SC00553"
 /note="SCF11.33c, unknown, len: 145 aa"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein SCF11.33c"
 /protein_id="CAB59610.1"
 /db_xref="GI:6137055"
 /translation="MRPFCTAEIRHTPPSPNRTVLPSPSEGVYTPAPPSSHPGAPA
 PAGRGAVVRLGVGVGVTVASGVSGSALVDRDGLGVRLVGEGLGVAVTVLDTVGRG
 GAGSGFCAVCATGAVGRVDPTTKWIVMTAVTLAAVQDSQMR"
 complement(3691..5388)
 /gene="SC00554"
 /note="synonyms: manA, SCF11.34c"
 complement(3691..5388)
 /gene="SC00554"
 /note="SCF73.01c, manA, probable secreted
 beta-mannosidase, len: 327 aa; similar to many e.g.
 MANA_STRLI P51529 mannan endo-1,4-beta-mannosidase
 precursor (EC 3.2.1.78) (363 aa), fasta scores: opt: 1189
 z-score: 1332.7 E(): 0, 61.1% identity in 288 aa overlap.
 Contains N-terminal signal sequence, and Pfam match to
 entry PF00150 cellulase, Cellulase (glycosyl hydrolase
 family 5). Also similar to SC2H4.16 (62.2% identity in 288
 aa overlap)

SCF11.34c, possible sugar hydrolase, partial CDS, len:
 >278 aa; C-terminus similar to SW:GUN4_THEFU (EMBL:U20093)
 Thermoanopora fusca endoglucanase E-4 precursor (EC
 3.2.1.4) (endo-1,4-beta-glucanase E-4), Celd, 880 aa;
 fasta scores: opt: 544 z-score: 555.0 E(): 1.5e-23; 42.8%
 identity in 215 aa overlap. Contains Pfam matches to
 entries PF00553 CBD_2, cellulose binding domain and
 PF00041 fn3, fibronectin type III domain. Overlaps with
 Streptomyces coelicolor ST573.01c, len: >327 aa;
 C-terminus similar to many e.g. MANA_STRLI P51529 mannan
 endo-1,4-beta-mannosidase precursor (EC 3.2.1.78) (363
 aa), fasta scores: opt: 1189 z-score: 1332.7 E(): 0, 61.1%
 identity in 288 aa overlap. Contains N-terminal signal
 sequence, and Pfam match to entry PF00150 cellulase,
 Cellulase (glycosyl hydrolase family 5)"
 /codon_start=1
 /transl_table=11
 /product="putative secreted beta-mannosidase"
 /protein_id="CAD55266.1"
 /db_xref="GI:24413729"
 /translation="MRPARLDTPRTLPAALLGLVGLVWVGCALCPGAALQSPAA
 RSAGVPAGAAQAGLHIGDGLRGNGDFVMRCVNHHTWYGETOSLADVRKALGN
 SVRVLSGDRHWSENGADYANVIEQCKANRLICVLEVHDTTGYAEADAAGTLDHAAD
 IWLKDLVLAGEQEDIVINIGNEPWGNTDPAGWTEPTVAQVKLRRAAGLQRTINVDAP
 NMGDWGVMRANARSYVDADPTGNLIFSIHMVSFDTAGTEITDYNAFVDAELPILI
 GEFGPADQYGDPEDETMATAEQLRLGYLAWSGNTPDVLIDLDFPSRLSGWGE
 RIFHGVNGIAQTSEATVYGDGAPDDTEATPAGAPSVTAVTSSVTLGWVAATDDTA
 VSGVDVVLVGDSTSVASSTTTATVTGLTAAATYTFVHARDACNRSARSATVEV
 TTGGGTPGGACSGVRYGVGEPWGFQGEIATVNRVTSNAGTLPWTLAFADAGQVTVNM
 WGGTATODAGAVSVAPASYYTATIPAGGTGTVLGTFTARKGDTNTAPAFRLNGCATTP"
 complement(3703..4005)
 /gene="SC00554"
 /note="Pfam match to entry PF00553 CBD_2, Cellulose
 binding domain, score 118.60, E-value 3e-33"
 complement(4054..4299)
 /gene="SC00554"
 /note="Pfam match to entry PF00041 fn3, Fibronectin type
 III domain, score 63.70, E-value 4.1e-15"
 complement(5394..5397)
 complement(5534..7129)
 /gene="SC00555"
 Query Match 40.3%; Score 903; DB 1; Length 314100;
 Best Local Similarity 65.0%; Pred. No. 1.2e-232;
 Matches 1386; Conservative 0; Mismatches 735; Indels 12; Gaps 3;
 QY 100 GGCACCAAAACAGGGATTGGTGGCCCAACATGCTCAACCTCGCATCTTACGCCACAT 159
 Db 13309 GCGCGGGAACCGCTCAGTGGTGGCGGAGCGGCTCAACCTGAAAGATCTCGCCCAAGAAC 13250
 QY 160 TCATCGGTATCGGACCAACAGCCCGGATTTTGACTATGCCGAAGAGTTAAGAAGCTA 219
 Db 13249 CCGCGCTCGCAACCCGCTGGGTGAGGAGTTCGACTACGCCGAGGCTTCGAGGCCCTC 13190
 QY 220 GATCTGCAGCGGTATAAAGGACCTGGCAGCGCTAATACAGATTACAGGACTGTGTGG 279
 Db 13189 GACCTCGGCGCGTAAACGGGACATCGCGAGGTGCTGACCACCTCGCAGGACTGTGTGG 13130
 QY 280 CCAGCAGATTACGCTCATTTATGGCCCTTCTTTATAGCATGCGCTGGCAGCGCCGCCG 339
 Db 13129 CCGCGCGCATTCGGCAACTACGGCCCGCTGATGATCGGAGTGGCTGGCAGCGCGGC 13070
 QY 340 ACCTACCGTATCGGTGATGCGCGTGGTGGGCGGTGGCTCCGGCTCACAGCGCTTCGCGCCCT 399
 Db 13069 ACCTACCGCATAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13010
 QY 400 CTCATAGCTGGCCAGACAAATGCCAATCTGGATAAAGCAGCGTTGCTTTTGGCCCATC 459
 Db 13009 CTCACAGCTGGCCGCAACCGGCAACCTGGACAAGCGCGCGCGCTGCTGTGGCGGTC 12950
 QY 460 AACAACAAATACGCTCCAAAANTCTCTGGCGGATCTAATGATCTACACAGGAACGTA 519
 Db 12949 AGAAGAAGTACGCCCAAGACCTCTCTCTGGCGGACCTCTCTCTGGTCTCACCGGCAAGTC 12890

QY 520 GCTCTGGAACATATGGGCTTTAAACTTTTGGTTTTCAGAGTGGCGAGACGATGTATGG 579
 Db 12889 GCGCTGGAGACGATGGGCTTCGAGACCTTCGCGGCTTCGCGGGGCGCGCGACGTGTGG 12830
 QY 580 GAGCCTGAAGAAGATCTATACCTGGGAGCAGAAACCGAATGCTCGGACACAGCGCTAT 639
 Db 12829 GAGCGCAGAGAGACGCTCTACTGGGCGCCGAGACCACTGGCTGACACCGCGGCTAC 12770
 QY 640 GAAGGTGACGAGAGCTCGAAATCCCTTGGGAGCGGTACAAATGGGACTCATCTATGTA 699
 Db 12769 ACCGGGACCGGAGCTGGAGAACCCTCGCGCGCTCCAGATGGTCTGATCTACGTC 12710
 QY 700 AACCCGGAAGGACCCAAACGCGAAGCCAGACCCCTATCGCTGCTCGCGGTGATATTCGTGAG 759
 Db 12709 AACCCGGAAGGCGCCAAACGCGAACCCTCGCGGACCGGATCGCGCGCGCGGACATCCGCGAG 12650
 QY 760 ACTTTTGGCGGAATGCAATGACGAAGAAACCGTGGCTCTCATACGGGTGGACAC 819
 Db 12649 ACCTTCGCGCGATGCGGATGAACGACGAGGAGACCGTCCGCTCATCGCGCGGCGACAC 12590
 QY 820 ACCTTCGGAACCCATCGTGTGCGGATGCGGAGAAATATGTGGGCGGAGCGCTGCC 879
 Db 12589 ACCTTCGGAAGACCCACGCGCGGCG---CCGGCGGACCGCTCGCGGACGCCGAG 12533
 QY 880 GCGCGAGGTATTGAAGAAATGACCTGGGGTGGAAACACACCTACGGCACCGGACACGCT 939
 Db 12532 GCGCGCGCATGGAGCAGCAGGCGCTCGGCTGGAAGAGCACCCACGCGACGGCGAAGGC 12473
 QY 940 GGGATACCATCACAGTGCAGTGAAGGCGCTGGACCAACGACCCCTACTCAATGAGC 999
 Db 12472 GGGACGCCATCACCTTCGCGCTTGGAGGTCACTTGGACGACGCGCCACCGAGTGGGG 12413
 QY 1000 AATAACTTTTTGAAACCTCTTTGGTTACGAGTGGGAGCTTACCAAAAGTCCAGCTGA 1059
 Db 12412 AACGGCTTCTCAAGAACCTCTTCGAGTTCGAGTACGAGCTGGAGCAGAGCGCGCGCG 12353
 QY 1060 GCTTATCATGGNAACCAAGAGCGTGGCGGCTGGCAACCATACCGGATGCAATGAT 1119
 Db 12352 GCGAACGATGGTTCGCAAGGACGACCGCGAG-----ATCATCCCGACGCGCACGAC 12299
 QY 1120 CCAGCAAGTCGACACCTCCATTTATGCTACCTACGAGCTCGCGCTGGCATGGACCTT 1179
 Db 12298 CCGGCCAAGACGACCGCCCGGAGTGTCAACACGACCTGCTGCGCTGGCGCTCGACCG 12239
 QY 1180 GATTACGAAAAATTTCTCGACGGTACTATGAAACCCCTGATGAGTTTGCAGATGCTTC 1239
 Db 12238 ATCTACGGCGCGATCTCCCGCGGTTCTACGAGAACCCGAGGAATTCGCGGACGCTTC 12179
 QY 1240 GCGAAGCATGGTACAACTGACACAGAGATATGGGACCAAGGTGCGCTACCTGGGA 1299
 Db 12178 GCGCGCGCTGTTCAAGCTGACCCACCGCGACATGGGCGCGAAGTCCCTGTACCTCGGC 12119
 QY 1300 CCAGAAGTGCCTCAGGAAGACCTCATCTGGCAAGACCTATACCAGATGTAAGCCATCCT 1359
 Db 12118 CCGGAGTCCCGAGGAGACCTTGATCTGGCAGGACCCGCTGCCGCGCGCGGAGCGGAG 12059
 QY 1360 CTGTAGAGAAACGATATTGAAGCCCTAAAGGCCAAATCTGGAATCGGAGCTGAGC 1419
 Db 12058 GTCATGACCGCGAGGACCTCGCCACCTCAAGACCAAGCTCTCGAGTCCGCGCTGTG 11999
 QY 1420 GTAAGCGAGCTGTTAAGCAGCGGATGCGGCTCTGATCTACTTTTGAAGAACTCTGACAAG 1479
 Db 11998 GTGTGCGAGCTGTGTGACCAACCGCGTGGGATCGGCTGCACTTCGCGGGAGCGACAAG 11939
 QY 1480 CCGCGGGTGGCAACGCTGACGATACGACTGCGCCCAACAAAGACTGGGAAGTAAC 1539
 Db 11938 CCGCGGGTGGCAACGCGCGCCGCTCCGCTGCAACCGGAGCGCGCTGGGAGGCTCAAC 11879
 QY 1540 AACCCCTCAGCAACTTCCCGAGGCTACTCAAAACACTAGAAGGTATCCAGGAGGACTTAAC 1599
 Db 11878 GAGCCGAGAGCTGGCGGAGGTGCTGCGGCTCTGGAGGTGTCCAGCGGAGTTCAC 11819

QY 1600 CAGGGCCCAATCAGATAACAAACAGATATCGTTGGCCGACCTGATTTGCTGGCGGCTGT 1659
 Db 11818 TCGGG---TCGGGCGCCAAAGAGTCTCCCTGGCGAGCTGATCTGCTGGCGGCTCG 11762
 QY 1660 GCGGGTGTAGAAAAAGCTCAAAAGATGCTGCCATAGGTGAGTGCAGTGCCTTTCAACCCG 1719
 Db 11761 GCGCGGTGAGAAAGCGCCCAAGGAGCGGCTTCCCGGTGGAGTGCCTTTTCGCGCG 11702
 QY 1720 GGACGAGCGGATGCCACCGCTGAGCAAAACCGATGTGGAAGCTTTTCAAGCACTAGAGCCA 1779
 Db 11701 GCGCGGTGAGCGGAGCGGAGGACACACCGAGCGCGAGTGTTCGAGGCCCTCGAACCG 11642
 QY 1780 GCGGCTGACGGCTTTAGAACTACATTAACCGGAGCATAAATATCCGCTGAGGAAATG 1839
 Db 11641 ACCGCGAGGGGTTCCGCAACTACCTTGGGCAAGGAAACCGGCTGCCCGCGAGTTCCTG 11582
 QY 1840 CTCGTAGACCGGGCGAGCTTCTGTCGCTTTCGGCACCAAGATGACTCTTTGGTAGGC 1899
 Db 11581 CTCCTCGACGGCGGAACCTGCTCACCTTGAGCGCGCGGAGATGACCTGCTTGGTGGC 11522
 QY 1900 GGTATCGTGTACTTGGGCGACCACTACGAGGTTTCGACGATGGAGTGTTCACAAATAAG 1959
 Db 11521 GCGCTCGGTGCTGGGCGGGGACACAGCAGTCCAGCTCGGCTGCTTACCCCGCAG 11462
 QY 1960 CCGGCTCAGCTATPCCAATGACTTCTTTTAAACCTGCTAGACTCAACACTAAATGGCGA 2019
 Db 11461 CCGGCTCCTGACCAACGACTTCTTCGTCACCTGCTGACCTGGCCACCACTGGAAG 11402
 QY 2020 GCCAGGATGAATCAGACAAAGTTTTTGAAGCAGAGACTTCAAACTGGCGAAGTAAG 2079
 Db 11401 TCGACCTCGAGGACCGGACCACTTTCGAGGGCGGCGCCGCCCGGAGGCTCAAG 11342
 QY 2080 TGSAGTGGCACCCGGGTAGACCTGATCTTCGGATCCAATTCAGAGTCAAGAGCCTCGCA 2139
 Db 11341 TSGGCGCGCAGCGCGCGACCTGGTCTTCGGCTCCACCGCGAGCTGGGCGCTGCGC 11282
 QY 2140 GAAGTGTACGGTGTGCGATCTGAAGAAAGTTTGTAAAGATTTGTGAAGGCGCTG 2199
 Db 11281 GAGGTACGCGAGCAGCAGCGGCGGAGAGATTCGTGACGACTTCGTGCGCGCTG 11222
 QY 2200 GCCAAGTATGAGACCTGGACCGGTTGATCTG 2232
 Db 11221 GTGAAGTCAATGAACCTCGACCTTCGACCTG 11189

RESULT 13 AE011760/c

LOCUS Xanthomonas axonopodis pv. citri str. 306, section 138 of 469 of the complete genome.
 ACCESSION AE011760 AE008923
 VERSION AE011760.1 GI:21107450

KEYWORDS Xanthomonas axonopodis pv. citri str. 306

SOURCE Xanthomonas axonopodis pv. citri str. 306

ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;

REFERENCE 1 (bases 1 to 11677)

AUTHORS

da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,

Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida

Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,

Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardoso, J.,

Chambergo, F., Ciapina, L.P., Cicarelli, R.M.B., Coutinho, L.L.,

Cursino-Santos, J.R., El-Dorfy, H., Faria, J.B., Ferreira, A.J.S.,

Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C.,

Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite

Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,

Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,

Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,

Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A.,

Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F.,

Takita, M.A., Tamura, R.E., Telxela, E.C., Tezza, R.I.D., Trindade dos

Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and

TITLE	Kitajima, J.P.
JOURNAL	Comparison of the genomes of two Xanthomonas pathogens with differing host specificities
MEDLINE	Nature 417 (6887), 459-463 (2002)
PUBMED	22022145
REFERENCE	12024217
AUTHORS	2 (bases 1 to 11677)
	da Silva, A.C.R., Ferro, J.A., Reinech, F.C., Farah, C.S., Furian, L.R., Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C., Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardoso, J., Chambergo, F., Ciapina, L.P., Ciccarelli, R.M.B., Coutinho, L.L., Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S., Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C., Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locati, E.C., Machado, M.A., Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J., Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M., Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A., Rossa, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F., Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and Kitajima, J.P.
TITLE	Direct Submission
JOURNAL	Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900, Brazil
FEATURES	Location/Qualifiers
source	1. .11677
	/organism="Xanthomonas axonopodis pv. citri str. 306"
	/mol_type="genomic DNA"
	/strain="306"
	/db_xref="taxon:190486"
	/note="pathovar: citri"
gene	103. .510
	/gene="rpls"
	/note="synonym: XAC1295"
CDS	103. .510
	/gene="rpls"
	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
	/codon_start=1
	/transl_table=11
	/product="50S ribosomal protein L19"
	/protein_id="AA036166.1"
	/db_xref="GI:21107451"
	/translation="MSKLNKYLILADFEAAQIQRLPEFNOGDTVVNVKVEGNREV QAYEGVIGTKNAGLNSFTYRKISHGFGVERVPTTHSAIDSVYKRRKVRAGKLY YIRLEGKARIKEDLAAAQAQAARQAARAAE"
gene	complement(781. .1353)
CDS	complement(781. .1353)
	/gene="XAC1296"
	/gene="XAC1296"
	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
	/codon_start=1
	/transl_table=11
	/product="conserved hypothetical protein"
	/protein_id="AA036167.1"
	/db_xref="GI:21107452"
	/translation="MAAQNVGAGLWSRVDAVARTGIAANFLMTGLLIVLIVDFRWTL ALFLPEGATLEVARHLNHIATWFLFGVSFVYVATGAVIPPLLIAPSLAGV RYPLANVLPHLGADAVVMWSPFISSACSMILSLAYIRWGWGRRKARMLTTPTHPAELAA AAEVPAHPASVPADPAPITDAPGGPRAPQ"
gene	complement(1357. .2265)
CDS	/gene="XAC1297"
	complement(1357. .2265)
	/gene="XAC1297"
	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
	/codon_start=1
	/transl_table=11
	/product="conserved hypothetical protein"
	/protein_id="AA036168.1"
	/db_xref="GI:21107453"
	/translation="MPKSAVLTGPIGNKLLLPALPMAGNIAQSLNGSLNATWICRY LGCEALTAANANSTIMPELIGSVFGIAGWASTILIGOAGARDIAQARKYGTSAFFFG GLSAVIATVGYLAPHLIIANGTPAASQALAEYLRVIFLAPMTLYLFAFLSALRGIT GSATPFRFLLSVLLDIIVFNPLLFGIPFPALGIAGANAWATLIIAOVVALAGLLYVL RKKRVAGLGRSDLRLEFRIDPAIRALIVRGVPMGLQVMISLISAMLSLVNGFGTGE TAAAFAGLQLWNTYQMPAMAVGAC"
gene	2647. .3342
CDS	/gene="XAC1298"
	2647. .3342
	/gene="XAC1298"
	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
	/codon_start=1
	/transl_table=11
	/product="transcriptional regulator"
	/protein_id="AA036169.1"
	/db_xref="GI:21107454"
	/translation="MVTNRATRLHLLDALRRRRPVAGAQAALRVSLRVLRYDRTAT LRAQADILGPGVGYVLRPGFLPALNFTDEDEALTIGRWVQADQPELAQAQAF AVARITATLPGTLRASLETSGLLVPADVAQPPAPLPILRGIRLEHVLWIDYADAE GAASORRVFPAMAFGLGCGYGIATWCELRGDFRFRADRVLADLADGERYPSRRHLII KRWRAATGIGADC"
gene	3416. .4057
CDS	/gene="gsta"
	3416. .4057
	/note="synonym: XAC1299"
	/gene="gsta"
	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
	/codon_start=1
	/transl_table=11
	/product="glutathione S-transferase"
	/protein_id="AA036170.1"
	/db_xref="GI:21107455"
	/translation="MLXTHPLSRGRVARMLEETGLPYRAVILDYGTMTKAPDYLA IN PMGKVPALRHGDTVVTENAAICAYLADLVPERHLAPPSPARADYVRLVFLTAGPVE ALLTAKESGVLPAARSAGYHEADLIQTLQAVAGREHLADGRFSAADLYMAAVGVGY TRTGMLEPRPFAFTAFALRHAARPAALRANATDDALINAHPNPDPRTVDITSS"
gene	4278. .4685
CDS	/gene="hslr"
	4278. .4685
	/note="synonym: XAC1300"
	/gene="hslr"
	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
	/codon_start=1
	/transl_table=11
	/product="heat shock protein 15 homolog"
	/protein_id="AA036171.1"
	/db_xref="GI:21107456"
	/translation="MNLEQAMTVRLDVMLWAARFTRSLAKNAVDTGKVDVAGOR PPSKALRSQGLRIDRGGIFETVVAALSDVRGPAPVAAQALYIEDSDSREARQWRQ QRAAERTGYRAPEKGPKDKRARRLTALGDIAL"
gene	complement(4809. .7079)
CDS	/gene="katG"
	complement(4809. .7079)
	/note="synonym: XAC1301"
	/gene="katG"
	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark; peroxidase"
	/codon_start=1
	/transl_table=11
	/product="catalase"
	/protein_id="AA036172.1"
	/db_xref="GI:21107457"
	/translation="MLLHTGIAMTTEAKCPFNHVVGTGTTNRDMPKOLRVDLLSOH SSKSNIPDPSFNADAFKHLDLOALKODLHMTDSDQDMPADFGYGHGFLFVYRMAHVS AGTNVIGDRGGCGGQRQAFAPLNSWPDNVSLKARLLMPIKQRTQAIISWADLMLG LVKDDDSQVPHTRDLEKLAQVNGLIYVNPPEGPDGPNPDIASARDIRDTFARMAND BETVALLAGHTTFCGTHGAPQVYGAPEAGELSQSGFGHWRHYSGGKGADTITSGI

gene	94	GGTGGCGGCACAAAACAGGGATTTGGTGGCCCAACATGCTCAACCTCGCATCTTTACGC 153
CDS	7013	GGTACCGGAACCAACATCGCATTTGGTGGCCCAACGACATGCGCGTGGATCTGCTCAGC 6954
gene	154	CAACATTCATCGCTATCGGACCCAAACGACCGCGGATTTTGACTATCCGAAGAGTTTAAG 213
CDS	6953	CACATTCGAGCAATCCACCCGCTGGACCCCTCGTTCAACTATGCCGACGCTTCAAG 6894
gene	214	AAGCTAGATCTGGCAGCGTTAAAGAGGACCTGGCAGCGCTAATGACAGATTCCACAGGAC 273
CDS	6893	CACCTGGACCTACAGCGCGCTCAAGCAGGACCTGCATCGCTTGATGACCGATTCCGAGGAC 6834
gene	274	TGGTGGCCACGACATACGCTCATATTGCCCCCTTTTATAGCATGGCGTGGCAGCAGC 333
CDS	6833	TGTTGGCGCGCCGACTTCGGCCATTACGGCCGCTGTTCTGTGCGCATGGCCTGGCATAGC 6774
gene	334	GCGCGGACCTTACCGTATCGGTGATGGCGGTGGCGGCTGGCTCCGCGCTCACAGCGGCTC 393
CDS	6773	GCGCGGACGTTATCGCATTGGCAGGACGTGGCGCGCGGACGTTGGCAGCAGCGCTTC 6714
gene	394	GCGGCTCTCAATAGCTGGCCAGACAATGCCAAATCTGGATAAAGACGCTTGTCTTTTGG 453
CDS	6713	GCCCGGCTCAACAGCTGGCGGACCAACGTCAGCCTGGACAAGGCGCGCGCTGCTGTGG 6654
gene	454	CCCATCAACAAATAACGCTCGAAAATCTCTCTGGCGGATCTAATGATATCTCACAGGA 513
CDS	6653	CCGATCAAGCAAAAGTACGCGCCAGGCCATTTCTTGGGCCGATCTGATGATCTTCCACG 6594
gene	514	AACGTAGCTCTGAAACTATGGGCTTTAAACCTTTTGGTTTTCAGGTGGCAGCAGCAGAT 573
CDS	6593	AAGTCGCGCTGGAATCGATGGGCTTGAAGACCTTCGGTTTTCGCGCGCGCGCGGAAGAT 6534
gene	574	GTATGGGAGCCTGAAGAAAGATGTATATCTGGGGAGCAGAAAACCGAATGGCTG---GGAGAC 630
CDS	6533	ACCTGGGAGCGGATCATGAGTCTGTACTGGCGCGCAACCAAGTGCCTGGGTGGGTGAC 6474
gene	631	AAGCGCTATGAAG----- 644
CDS	6473	GAGCGCTATTCTGGCGCGGTTGCCCGCGGCTGGAGCAAGCGACGCGGGTGTGTTGAAGGAC 6414

1361	GAGCCGAGGAGCGTCTACTGGGCCCGGAGACCACCTTGGCTCGACGACCGCGGTAC	1420	
Qy	640	GAAGGTGACGAGAGCTCGAAAAATCCCCTGGGAGCGGTACAAAATGGAGCATCACTATTGTA	699
Db	1421	ACCGCGCCCGCAGCTGGAGAACCCGCTCGGCGCGTCCAGATGGTGTGATCTACGTC	1480
Qy	700	AACCCGAAAGGACCCAAACGGCAAAGCAGACCCCTATCGCTGCTCGGCGTGATATTCGTGAG	759
Db	1481	AACCCGAGGCGCCCAACGCAACCGGAGCCCATCGCGCCCGCGCACATCCCGAG	1540
Qy	760	ACTTTTGGCCGAATGCAANTGAATGACGAGAACCTGCTCTCATACGGGTGGACAC	819
Db	1541	ACCTTCGCGCATGGGCATGAACAGCAGGAGACCGCTCGCTCATPCGCGCGGACAC	1600
Qy	820	ACCTTCGAAAAACCCATGTGCTGGCGATGCGGAGAAATATGTGGGCCGAGACCTGCC	879
Db	1601	ACCTTCGGCAAGACCCACGCGCGGGC--CCGGCCGACCCCTCGGGACGACCCCGAG	1657
Qy	880	CGCGCAGGTATTGAAGAAATGACCTTGCGGTGGAAACAACCTACGCGCACCGGACACGGT	939
Db	1658	CGCGCGCCCATGGAGCAGCAGGCGCTCGCTGGAAGAGCACTACGCGCACGGCAAGGC	1717
Qy	940	CGSGATACCATCACCTAGTGACATAGAGGCGCTGGACCAAGACCCCTACTCAATGGAC	999
Db	1718	GGCAGGCCATACCTTCGCGCCTGGAGGTCACTTGGACAGCAGCCACCCAGTGGGC	1777
Qy	1000	AATAACTTTTTGAAAAACCTCTTTGGTTACAGTGGGAGCTTACCAAAAAGTCAGGTGGA	1059
Db	1778	AACGGCTTCTCAAGAACCTCTTCGAGTTCGAGTACGAGCTGGAGCAGACCCGCGCGC	1837
Qy	1060	GCTTATCAGTGGAAACCAAAAGACGGTGGCGGGCTGGCACCATACCGGATGCACATGAT	1119
Db	1838	GCCAACAGTGGGTGCGCAAGAGCAGCACCGGAG-----ATCATCCCGACGCGCACGAC	1891
Qy	1120	CCGAGCAAGTCGACGCTCCATTATTGCTCACTACGAGCTGGCGCTGGCCATGGACCTT	1179
Db	1892	CCGGCCAAAGACACCGCCCCAGATGCTCACCACGACCTGTGCTGCGCCCTCGACCG	1951
Qy	1180	GATTACGAAAAAATTTCTCGACGCTACTATGAAAACCTTGATGAGTTTCAGATGCTTTC	1239
Db	1952	ATCTACGCGCGATCTCCCGCGGTCTACGAGAACCGGAGGAATTCGCGGACGCTTC	2011
Qy	1240	CGGAAGCATGTTACAACTTGACACACAGAGATATGGGACCAAGGTGCGCTACCTGGGA	1399
Db	2012	CGCGCGCCTGGTTCAAGCTGACCCACCGCGACATGGGCCGAAGTCCCTGTACCTCGGC	2071
Qy	1300	CCAGAAGTGCCTCAGGAAGACCTCATCTGCGAAGACCCCTATACCAGATGTAAGCCATCCT	1359
Db	2072	CCGGAGTCCCCGAGGAGACCTTGATCTGCGAGGACCCGCTGCGCGAGCGGCGAG	2131
Qy	1360	CTTTGTAGCAAAACCATATTTGAAGGCTTAAAAGCCAAAATCTTGAATCGGACGTGACG	1419
Db	2132	GTCAATCGACGCGGAGAGCTCGCACCCCTCAAGCAAGTCTCTCGAGTCCGCGCTGTGCG	2191
Qy	1420	GTAAGCAGTGGTAGCAGGATGGGCTTCTGCACTACTTTTTAGAAACTCTGACAAG	1479
Db	2192	GTGTGCGAGCTGTGACCAACCGGTGGGCATCGCGCTCCACCTTCGCGGGAGCGACAAG	2251
Qy	1480	CGGCGCGTCCCAACGTTGCACTATACGACTGGCCCCACAAAAGACTGGGAAGTAAAC	1539
Db	2252	CGGCGCGGTGCCAACGCGGCCGATCCGCTCGAACCAGCAGCGCGCTGGGAGGTCAAC	2311
Qy	1540	AACCTCAGCAACTTGCAGGGTACTCAAAACACTAGAAGGTATCCAGGAGGACTTTAAC	1599
Db	2312	GAGCCGACAGCTGGCGAGGTGCTCGCGCTCTTGAGGGGTGTCCAGCGGAGTTC AAC	2371
Qy	1600	CAGCGCAATCAGATAACAAAGCAGTATCGTTGGCCGACCTGATGTGCTGGCGCGGTGT	1659
Db	2372	TCCGGG---TCGCGCGCAAGAAGTCTCCTGGCGCACCTGATCGTCTCGCGCGGTGCG	2428
Qy	1660	CGGCGTGTAGAAAAGCTGCAAAAGATGCTTGGCCATGAGGTGCAAGTGCCTTTCAACCGG	1719
Db	2429	CGCGCGTTCAGAAAGCGCCAGAGAGCGGGTTCGCGTGGAGGTTCCTTCCTTCGCGCGG	2488

Qy	1720	GGACGAGCGGATGCCACCGCTGTGAGCAAAACGATGTGGAAAGCTTTTCGAAGCACTAGAGCCCA	1779
Db	2489	GGCCGCGTGGACGCGAGCGGAGGAGCACACGAGCCGAGTCTGGTGTAGCTTCCTCGAACC	2548
Qy	1780	CGGCTCAGCGCTTTTAAAACTACATTAACCCGNGCATAAAGTATCCGCTGAGGAATG	1839
Db	2549	ACGCCGACGGGTTCGCCAACTACCTGGGCAAGGAAACCGGCTGCCCGCGGAGTTCCCTG	2608
Qy	1840	CTCGTAGACGGGGCGGACGCTTCTGTCTCTTTTGGCACCCAGAAATGACTGCTTTGTAGGC	1899
Db	2609	CTCTCGACCGGGCAACCTGCTCACCTTGAGCGGCCCGGAGATGACCGTCTGCTGGC	2668
Qy	1900	GGTATGGGTACTTGGGCACCACTACGACGGTTCGACGATGGAGTGTTCACAAATAG	1959
Db	2669	GGCTGGCGGCTGTGGCGGGGACACACGACGTCCTGAGCTCGGCGTCTTCACCCGCGACG	2728
Qy	1960	CCGGTTCAGCTATCCAAATGACTTCTTTGTTAAACCTGTAGACCTCAACACTAAATGGCGA	2019
Db	2729	CCGGCTCGCTGACCACGACTTCTGCTCACTGCTCGACCTTGGGCACCACTGGAAG	2788
Qy	2020	GCCAGCGATGAATCAGACAAAGTTTGTGAAGGCGAGAGACTTCAAACTGGCGAAGTAAAG	2079
Db	2789	TCGACGTCGGAGGACCGACACCTTCGAGGGCGGACGCGCCACCGCGAGGTCAAG	2848
Qy	2080	TGGAGTGGCACCCGGGTAGACCTGATCTTGGATCCAAATTCGAGCTTAAGAGCCCTCGCA	2139
Db	2849	TGGCGCGGACGCGCGCGACCTGGTCTTCGGCTTCCAAACCGGAGCTGCGGCGCTCGCC	2908
Qy	2140	GAAGTGTACGGCTGTGAGATTCCTGAAGAAAAGTTGTTAAAGATTTTGTGAAGGCGCTGG	2199
Db	2909	GAGTCTACGGACGACGACGCGGGCGGAGAGTTCTGTCAGCACTTCGTCGCGCGCTGG	2968
Qy	2200	GCCAAAGTAATGGACCTGGACCGGTTTGATCTG	2232
Db	2969	GTGAAGGTTCATGAACCTCGACCGTTTCGACCTG	3001
RESULT 15	AP005937	298900 bp	DNA linear BCT 28-MAR-2003
LOCUS	Bradyrhizobium japonicum USDA 110	DNA, complete genome, section 3/31.	
DEFINITION	AP005937 GI:27348808		
ACCESSION	AP005937	BA000040	
VERSION	Bradyrhizobium japonicum USDA 110		
KEYWORDS	Bradyrhizobium japonicum USDA 110		
SOURCE	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Bradyrhizobiaceae; Bradyrhizobium.		
REFERENCE	1		
AUTHORS	Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., Watanabe, A., Idesawa, K., Iriguchi, M., Kawashima, K., Kohara, M., Matsumoto, M., Shimpo, S., Tsuruoka, H., Wada, T., Yamada, M. and Tabata, S.		
TITLE	Complete genomic sequence of nitrogen-fixing symbiotic bacterium Bradyrhizobium japonicum USDA110		
JOURNAL	DNA Res. 9 (6), 189-197 (2002)		
MEDLINE	22484998		
PUBMED	12597275		
REFERENCE	2		
AUTHORS	Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., Watanabe, A., Idesawa, K., Iriguchi, M., Kawashima, K., Kohara, M., Matsumoto, M., Shimpo, S., Tsuruoka, H., Wada, T., Yamada, M. and Tabata, S.		
TITLE	Complete genomic sequence of nitrogen-fixing symbiotic bacterium Bradyrhizobium japonicum USDA110 (supplement)		
JOURNAL	DNA Res. 9 (6), 225-256 (2002)		
MEDLINE	22485002		
PUBMED	12597279		
REFERENCE	3 (bases 1 to 298900)		
AUTHORS	Kaneko, T.		
TITLE	Direct Submission		


```

/codon_start=1
/transl_table=11
/product="glycerol-3-phosphate dehydrogenase [NAD(P)+]"
/protein_id="BAC45833.1"
/db_xref="GI:27348817"
/translation="MTAFHSVAVIGAGAGWTALAMVAARAGRSVTLWNAEHATRIA
STRDPRPLGVRLADIVVTSELALAAADMLLIATPAHLRGAVNMLASHIARETPV
VACAKGISHGTHKFTVEIAEALHAQPAILSGFSFADVDVARGLPATVTLAAKEELA
SLVQALGSPFRPHSTDVRGVEIGGAAKNVLAIAAGIVVGRNIGASALAAITRRGF
SELAIRLGRACARPETLSGLSGLDLSCSTAOSRNFALGIALGRGAAPAGKLAEG
AFTAPVLVELAAARNVMPVSQAVAAIILDKLITIDAAIEGLITRPFRAEE"
9344..9790
/gene="blr0569"
CDS
9344..9790
/gene="blr0569"
/note="ORF_id:blr0569
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAC45834.1"
/db_xref="GI:27348818"

Query Match      39.8%;  Score 890.8;  DB 1;  Length 298900;
Best Local Similarity 65.5%;  Pred. No. 2.5e-229;
Matches 1405;  Conservative 0;  Mismatches 717;  Indels 24;  Gaps 6;

QY  89  GTGCAAGTGGCGGCACAAAACACAGGATTTGGTGGCCCAACATGCTCAACCTCGGCATCT 148
DB  231652  GTGGAAAACGGCTGCGCGCACCGCGATTTGGTGGCCCAACCGCATCGACATCGAGATGC 231711
QY  149  TACGCACATTCATCGTATCGGACCCAAACGACCCGGATTTTCACATATGCCGAGAGT 208
DB  231712  TCACAAGAAATTCGGGCTGTCCGATCCGATGGGCGAGGAATTCGACTACGCCAAGGAAT 231771
QY  209  TTAAGAAGCTAGATCTGGCAGCGGTTAAAGAGACCTGGCAGCGCTAATGACAGATTCAC 268
DB  231772  TCAAGACGCTGACCTGAACGGGTGCATCAAGGACCTGACCGCTCTGATGACGGACTCGC 231831
QY  269  AGGACTGGTGGCCACGACATACGCTCATATTATGGCCCTTCTTTATGACGATGGCGTGC 328
DB  231832  AGGATGGTGGCCCGCCCTTCGTCATATTATGGCGGCTGTATGATCCGCAATGGGCTGSC 231891
QY  329  ACAGCGCGGCACTACCGTATCGGTGATGGCGGTGGCGGTGGCTCGGCTCACAGC 388
DB  231892  ACAGCGCGGCACTATCGCATCACCGATGGCGGTGGCGGCGCGCGGTGCGGTTCAGCAGC 231951
QY  389  GCTTCGCGCTCTCAATAGCTGGCCAGACAATGCCAATCTGGATTAAGCACGCTTGCTTC 448
DB  231952  GTTTCGCGCGCTCAACAGCTGGCCGACACGCGCAACCTCGACAAGCGCGCGCGCTGCG 232011
QY  449  TTTGGCCCATCAACAAATAACGGTCGAAAATCTCTGGCGGATCTAATGATACTCA 508
DB  232012  TCTGGCCCATCAAGCAGAAATACGGCCCAAGATCTCTGGCGGATTTGATGGTGTCTCG 232071
QY  509  CAGGAACAGTACTCTGGAAACTATGGCTTTAAAACTTTTGGTTTTCAGGTGGCAGAG 568
DB  232072  CCGGCAATGTCCGCTGGAGTCGATGGGCTTCAAGACGTTTGGTTTCGCGGTGGCCGCG 232131
QY  569  CAGATCTATGGAGGCTGAAGAAGATGTATCTGGGGAGCAGAAACCGAATGGCTGGGAG 628
DB  232132  CCGAGCTCTGGAGCC---GGNAGAGCTCTATTGGGTTCCCGAAGGCACCTGGCTGGGCG 232188
QY  629  ACAAGCGTATGAAGTGAACGAGGCTCGAAAATCCCTTGGAGCCGTTACAAATGGGAC 688
DB  232189  ATGAGCGGTACAGCGCGAGCCAGCTCGCGGAGCGCTCGGCGGCTGCGAGATGGGCC 232248
QY  689  TCATCTATGTAACCCGAGAGACCCACGGCAAGCCAGACCTATCGCTGTGCGCGGTG 748
DB  232249  TCATCTACGTCAATCCGAAAGGCCGAACGGCAAGCCGACCGCATCGCGCGGCAAAAGG 232308
QY  749  ATATTCTGTGAGACTTTTGGCCGAATGGCAATGAATGACGAAGAAACCGTGGCTCTCATAG 808
DB  232309  ACATCCGGAGACCTTCTTCCGATGGCGATGAACGAGGAAACCGTGGCGCTGATGCG 232368

```

```

QY  809  CGGGTGGACACACACTTTCGGAAAAAACCCATGTGTGCGGATGCGGAGAAAAATATGTGGCC 868
DB  232369  CGGGCGGTTCATACCTTCGCAAGACCCATGTGTGCGGCGATCCGTGCG---CTGTGCGGAC 232425
QY  869  GAGACCTTGGCCGCCAGGTATTGAAGAAATGAGCCTGGGGTGGAAAAACACCTACGGCA 928
DB  232426  CGGACCGGAAGCGGGCGCACTCGAGGATCGAGGGCTCGGGTGGAGAGCAAGACGCTT 232485
QY  929  CCGGACACGGTGGCGATACCATCACAGTGGACTAGAAGCGCCTGGGACCAAGACCCCTA 988
DB  232486  CGGGCTTGGCGGCGATTCCATCACAGCGGACTCGAGGTGACATGGACGACGACGCGA 232545
QY  989  CTCATATGAGCAATTAACCTTTTGAAGAACCTCTTTGGTTTACAGTGGGAGCTTACCAAAA 1048
DB  232546  CGAAGTGGAGCAACAACCTTCTTCGAGAACCTTCTCAAGTACGAATGGAGCTGACGAAGA 232605
QY  1049  GTCCAGCTGGAGCTTATCAGTGGAAAAACCAAGAGCGTGCCTGGGGCTGGCACCACATCCGG 1108
DB  232606  GCCCGGGCGGCGGAACCAAGTGGACGGCGCAAGGGTGCAGGCG-----ATCATTCGCG 232659
QY  1109  ATGCACATGATCCACAGCAAGTCGCACGCTCCATTTATGCTCACTACGACCTGGCGCTGC 1168
DB  232660  AGCCCTTCGACAAGTCGAAGAGCACCGCGCGACGATGCTGACACCGACCTCTCGCTGC 232719
QY  1169  GCATGGACCTGATTACGAAAAAATTTCTGACGGTACTATGAAAAACCTTGATGATTTG 1228
DB  232720  GCCTCGATCCGCGCTTATGAGAAAGATCTCGCGCGCTTCTTGGAGAACCCGGATCAGTTCG 232779
QY  1229  CAGATGCTTTCGGAAGACNTGTGTAACACTGACACACAGAGATATGGGACCAAGGTGC 1288
DB  232780  CGGATGCAATTCGCCCGCGCTGGTTCAAGCTCACCCATCGGCACATGGGTCCGATCCACG 232839
QY  1289  GCTACTCTGGGACCAAGATGCTCAGGAAGACCTCATCTGGCAAGACCCATATACACAGATG 1348
DB  232840  GCTATCTCGGCGCGTGTGTCGCGAAGGAGACGCTGATCTGGCAGGATCCGATCCCGCGCG 232899
QY  1349  TAAGCCATCTCTTTAGACGAAAAAGATATTGAAGGCTTAAAGCGCTTAAAGCCAAATCTTGG 1408
DB  232900  TGAATCACAGAGTGTGTGACGATCAGGACATCGCTGCGCTGAAGACGAAGATCTCTGGCT 232959
QY  1409  CGGGAGTACGCTAGCAGAGCTGTTAAGCACGGCATGGCTTCTGCATCTACTTTTAGAA 1468
DB  232960  CGGGCTCTCTGGTCCGCGAGCTGGTCTCGACCGCATGGGCTGGCGCTCGAGCTTCCGCG 233019
QY  1469  ACTCTGACAAGCGCGCGGTGCCAACGCTGACGCTATAGACTGGCCCCACAAAAAGACT 1528
DB  233020  GTTTCGGACAAGCGCGCGCGCAACCGCGCGCATTCGTCTCAGCCCGCAGAGGACT 233079
QY  1529  GGGAGATTAACAACCTCTAGCAACTTTGCCAGGGTACTCAAAACACTAGAAAGTATCCAGG 1588
DB  233080  GGGAGGTAAACACCGCGGCTCAGCTCTCGAAGGTGCTCGGAAGCTCGGAAGCGATCCAGA 233139
QY  1589  AGGACTTTAACCCAGCGCAATCAGATAACAGCAGTATCGTTGGCCGACCTGATGTGCG 1648
DB  233140  AGGAGTTCAACGCTTCGGCGCGCGGAGAA---GGTCTCGCTCGCGGACCTGATCGCTCC 233196
QY  1649  TGGCCGCTGTGCGGGTGTAGAAAAAGCTGCAAAAAGATGCTGGCCATGAGGTGCAAGGTGC 1708
DB  233197  TCGGGCGGACCCCGCGGTGAGAAAGCGCGCGAAGGACGCGCGCTCGACGTCAAGGTGCG 233256
QY  1709  CTTTCAACCCCGGACGAGCGGATGCCACCGTGTGAGCAACCAAGATGGAAGCTTTCGAAG 1768
DB  233257  CTTTCAACCGGCGCGCATGGACGCTCGCAGGAGCAGACGATCGGCTTCTCTCGCGC 233316
QY  1769  CACTAGAGCAGCGCTGACGGCTTTAGAACTACATT---AAACCGGAGCATAAAGTAT 1825
DB  233317  CGCTGAGCGCGCGCGCATGTTGTTTCCGCACTAGCTGGCAAGCGGACGAGTCTCTGCG 233376
QY  1826  CCGCTGTAGAAATGCTCTAGACCGGCGCGCTTCTGCTGCTTTTGGCCACCAAGATCA 1885
DB  233377  AGCAGGAAGAGCGCTCTGTCGATCGGCACAGCTTCTTGAAGCTCACCGGACCCGAGTTGA 233436

```

```

QY 1886 CTGCTTTAGCGGTATCGCTACTGGGCACCACTACGACGGTTTCGACGATGGAG 1945
Db 233437 CCGTCTGTGCGCGGTCTGCGCGTCTCGGTGCCAATGCGAACGGTTTCGAAGCACGGCG 233496

QY 1946 TGTTTACAAATFANGCCGGTCCAGCTATCCAAATGACTTCTTTGTAAACCTGCTAGACCTCA 2005
Db 233497 TCCTCACCGCGAAGGTGGTGGCTGACCAAGACTTCTTCGTCAACCTGCTCGACATGA 233556

QY 2006 ACACATAATGGGAGCCAGCCGATGAATCAGACAAAGTTTTTTGAAGGCGAGACTTCAAAA 2065
Db 233557 GCACGCAGTGGGCGCC-----GGCGGCCGACGCGCCTACGAGGCTCGCGACCGCAAGA 233610

QY 2066 CTGGCGAAGTAAAGTGGAGTGGCAACCGGGTAGACCTGATCTTCGGATCCAAATCCGAGC 2125
Db 233611 CCAAGCGGTGAAGTGGACCGGCGACGCGCGCGATCTCATCTTCGGCGCGCACTCGCAGC 233670

QY 2126 TAAGAGCCCTCGCAGAAAGTGTACGGCTGTGCGAGATTCTGAAGAAAAGTTTGTAAAGATT 2185
Db 233671 TCCGCGCTACGCCGAGGTCTATGCCACCTCGGACTCCAAGGAGCAGTTCGTGAAGGACT 233730

QY 2186 TTGTGAAGGCTGGGCCAAGTAATGGACCTGGACCGGTTTGATCT 2231
Db 233731 TCGCCAAGGCTGGACCAAGGTGATGAACCTCGACCGGTTGATCT 233776

```

Search completed: October 8, 2003, 12:53:49
Job time : 8128.46 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2003, 06:24:20 ; Search time 575.415 Seconds
(without alignments)
10499.111 Million cell updates/sec

Title: US-09-884-889-7
Perfect score: 2238
Sequence: 1 atggaatacacaacacac.....accggttgatctgaataa 2238

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2238	100.0	2238	19 AAV06555	Microscilla furves
2	2238	100.0	2238	24 ABS53945	DNA encoding Micro
3	844.4	37.7	2331	18 AAT90400	Mycobacterium tube
4	844.4	37.7	2331	20 AAX00817	M. tuberculosis ca
5	844.4	37.7	82993	24 ABX09140	Mycobacterium tube
6	844.4	37.7	4403765	22 AAI99683	Mycobacterium tube
7	844.4	37.7	4411529	22 AAI99682	Mycobacterium tube
8	842.8	37.7	2331	20 AAX00818	M. tuberculosis ca

9	832.4	37.2	2235	19 AAT89967	Mycobacterium tube
10	721.8	32.3	2221	19 AAT89968	Mycobacterium tube
11	721.8	32.3	4795	14 AAO51531	M.tuberculosis H37
12	721.8	32.3	4795	20 AAX16969	M.tuberculosis kat
13	721.6	32.2	2196	11 AAO6815	Sequence encoding
14	672.2	30.0	2262	19 AAV05554	Alcaligenes (Deley
15	672.2	30.0	2262	24 ABS53944	DNA encoding Alcal
16	591	26.4	2839	24 ABQ78599	Nucleotide sequenc
17	570.6	25.5	1852	21 AAF07490	Fusarium venenatum
18	357.6	16.0	1489	21 AAZ36101	Nucleic acid sequ
19	345	15.4	5059	23 AAS89244	DNA encoding novel
20	279.2	12.5	660	21 AAA49864	Mycobacterium tube
21	270.2	12.1	671	21 AAF14551	Aspergillus oryzae
22	258.6	11.6	824	23 AAS73010	DNA encoding novel
23	251.8	11.3	620	17 AAT29131	katG gene fragment
24	250.2	11.2	620	17 AAT29129	katG gene fragment
25	250.2	11.2	620	17 AAT29132	katG gene fragment
26	248.6	11.1	620	17 AAT29130	katG gene fragment
27	190.4	8.5	4653	23 AAS73484	DNA encoding novel
28	190.4	8.5	4653	23 AAS89265	DNA encoding novel
29	190.4	8.5	4653	23 AAS89808	DNA encoding novel
30	190.4	8.5	4653	23 AAS90173	DNA encoding novel
31	163.4	7.3	6240	23 AAS90138	DNA encoding novel
32	157.2	7.0	391	19 AAV70424	DNA sequence of ka
33	157.2	7.0	391	24 ABL46034	Mycobacterium tube
34	155.6	7.0	391	19 AAV70426	DNA sequence of ka
35	155.6	7.0	391	19 AAV70427	DNA sequence of ka
36	155.6	7.0	391	24 ABL46036	Mycobacterium tube
37	155.6	7.0	391	24 ABL46037	Mycobacterium tube
38	154	6.9	391	19 AAV70425	DNA sequence of ka
39	154	6.9	391	24 ABL46035	Mycobacterium tube
40	150.2	6.7	522	24 ABQ78598	Fragment of Agroba
41	127.6	5.7	538	24 ABQ23234	Oligonucleotide fo
42	127.6	5.7	538	24 ABQ23235	Oligonucleotide fo
43	124.8	5.6	763	24 ABQ21416	Oligonucleotide fo
44	124.8	5.6	763	24 ABQ21417	Oligonucleotide fo
45	108.6	4.9	763	24 ABQ21414	Oligonucleotide fo

ALIGNMENTS

RESULT 1	AAV06555	AAV06555 standard; DNA; 2238 BP.
ID	AAV06555	AAV06555 standard; DNA; 2238 BP.
XX	AC	AAV06555;
XX	DT	03-JUL-1998 (first entry)
XX	DE	Microscilla furvescens catalase-53CA1 gene.
XX	KW	Catalase; epoxidation; hydroxylation; biosensor; paper bleaching;
XX	KW	pasteurisation; ss.
XX	OS	Microscilla furvescens.
XX	XX	Key Location/Qualifiers
XX	FT	1..2238
XX	FT	/*tag= a
XX	FT	/product= "Catalase-53CA1"
XX	PN	WO9800526-A1.
XX	PD	08-JAN-1998.
XX	PF	03-JUL-1997; 97WO-US16513.
XX	PR	03-JUL-1996; 96US-0674887.
XX	PA	(RECO-) RECOMBINANT BIOCATALYSIS INC.
XX	PI	Adhikary RS, Robertson DE, Sanyal I;

XX DR WPI; 1998-086953/08.
XX DR P-ESDB; AAW33810.
XX PT New bacterial catalases, related nucleic acid vectors and
XX PT transformed cells - used as oxidising agents and for detecting or
XX PT destroying hydrogen peroxide, e.g. in biosensors
XX PS Claim 3; Fig 2; 35pp; English.
XX CC The present sequence is of the Microscilla furvescens catalase-53Cal
XX CC gene. Fragments of the gene can be used to identify related sequences.
XX CC Catalase-53Cal may be used to catalyse oxidation reactions such as
XX CC epoxidation or hydroxylation. The enzyme can also be used to detect or
XX CC destroy hydrogen peroxide, e.g. in connection with glyoxylic acid
XX CC production, biosensors, contact lens cleaning, pulp/paper bleaching and
XX CC pasteurisation of dairy products. Antibodies raised against
XX CC catalase-53Cal can be used to screen libraries for detection and
XX CC purification of cells containing the enzyme.
XX SQ Sequence 2238 BP; 634 A; 545 C; 605 G; 454 T; 0 other;
Query Match 100.0%; Score 2238; DB 19; Length 2238;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGGAATACAAACACTCAGGATCTTACGTATAACAAACACTGGCGGAAATGC 60
Db 1 ATGGAATACAAACACTCAGGATCTTACGTATAACAAACACTGGCGGAAATGC 60
Qy 61 CCTTTTACCGAGGTTTCGCTTAAGCAAGTGCAGGTGGCGGCACCAAAACAGGATGG 120
Db 61 CCTTTTACCGAGGTTTCGCTTAAGCAAGTGCAGGTGGCGGCACCAAAACAGGATGG 120
Qy 121 TGGCCCAACATGCTCAACCTCGGATCTTACGCCAACATTCATCGCTATCGGACCCAAAC 180
Db 121 TGGCCCAACATGCTCAACCTCGGATCTTACGCCAACATTCATCGCTATCGGACCCAAAC 180
Qy 181 GACCCGGATTTGACTATGCCGAAGAGTTTAAAGAGTAGATCTGGCAGCGTTTAAAG 240
Db 181 GACCCGGATTTGACTATGCCGAAGAGTTTAAAGAGTAGATCTGGCAGCGTTTAAAG 240
Qy 241 GACCTGGCAGGCTTAATGACAGATTACAGGACTGGTGGCAGCAGATTACGGTCATTAT 300
Db 241 GACCTGGCAGGCTTAATGACAGATTACAGGACTGGTGGCAGCAGATTACGGTCATTAT 300
Qy 301 GGCCCTTCTTTATACGATGGCGTGGCAGCGCGGCACTACCGTATCGGTGATGGC 360
Db 301 GGCCCTTCTTTATACGATGGCGTGGCAGCGCGGCACTACCGTATCGGTGATGGC 360
Qy 361 CGTGGTGGCGTGGCTCCGGCTCACAGCGTTCGCGCTCTCAATAGCTGGCCAGCAAT 420
Db 361 CGTGGTGGCGTGGCTCCGGCTCACAGCGTTCGCGCTCTCAATAGCTGGCCAGCAAT 420
Qy 421 GCCAATCTGGTAAAGCAGCGTTCCTTTGGCCCATCAACAAAANTACGGTCGAAA 480
Db 421 GCCAATCTGGTAAAGCAGCGTTCCTTTGGCCCATCAACAAAANTACGGTCGAAA 480
Qy 481 ATCTCTGGCGGATCTAATGATACTCACAGGAAACGTAGCTCTGGAACCTATGGGCTTT 540
Db 481 ATCTCTGGCGGATCTAATGATACTCACAGGAAACGTAGCTCTGGAACCTATGGGCTTT 540
Qy 541 AAAAATTTTGGTTTTCAGGTGGCAGCAGATGATGGAGCCTCAAGAGATGTATAC 600
Db 541 AAAAATTTTGGTTTTCAGGTGGCAGCAGATGATGGAGCCTCAAGAGATGTATAC 600
Qy 601 TGGGAGCAGAAACCGAATGGCTGGGAGACAAAGCGCTATGAGGTGACGAGAGCTCGAA 660
Db 601 TGGGAGCAGAAACCGAATGGCTGGGAGACAAAGCGCTATGAGGTGACGAGAGCTCGAA 660
Qy 661 AATCCCTGGAGCGCTACAAATGGGACTCATCTATGTAACCCCGGAGGACCCAAACGGC 720
Db 661 AATCCCTGGAGCGCTACAAATGGGACTCATCTATGTAACCCCGGAGGACCCAAACGGC 720

Qy 721 AAGCCAGACCCCTATCGCTGCTGCGTGATATCGTGAGACTTTTGGCCGAATGCAATG 780
Db 721 AAGCCAGACCCCTATCGCTGCTGCGTGATATCGTGAGACTTTTGGCCGAATGCAATG 780
Qy 781 AATGACGAAGAAACCGTGGCTCTCATAGCGGGTGACACACCTTCGGAAAAACCATGGT 840
Db 781 AATGACGAAGAAACCGTGGCTCTCATAGCGGGTGACACACCTTCGGAAAAACCATGGT 840
Qy 841 GCTCCGATCGGAGAAATATGTGGCCGAGAGCTGCCCGCAGGTATTTGAAGAAATG 900
Db 841 GCTCCGATCGGAGAAATATGTGGCCGAGAGCTGCCCGCAGGTATTTGAAGAAATG 900
Qy 901 AGCCTGGGGTGGAAAAACACCTACGGCACCGGACACCGTGGGATACCATCACCAGTGA 960
Db 901 AGCCTGGGGTGGAAAAACACCTACGGCACCGGACACCGTGGGATACCATCACCAGTGA 960
Qy 961 CTAGAAGGCGCTGGACCAAGACCCCTACTCAATGGAGCAATACTTTTGAAGAACTC 1020
Db 961 CTAGAAGGCGCTGGACCAAGACCCCTACTCAATGGAGCAATACTTTTGAAGAACTC 1020
Qy 1021 TTTGGTTACGAGTGGAGCTTACCAAAAGTCCAGCTGAGCTTATCAGTGGAAACCAAAA 1080
Db 1021 TTTGGTTACGAGTGGAGCTTACCAAAAGTCCAGCTGAGCTTATCAGTGGAAACCAAAA 1080
Qy 1081 GACGGTCCCGGGCTGGCACCATACCGGATGCACATGATCCAGCAAGTCCAGCTCCA 1140
Db 1081 GACGGTCCCGGGCTGGCACCATACCGGATGCACATGATCCAGCAAGTCCAGCTCCA 1140
Qy 1141 TTTATGCTACTACGAGCTGGCGCTGGCGATGACCGCTGATTAACGAAAAATTTCTCGA 1200
Db 1141 TTTATGCTACTACGAGCTGGCGCTGGCGATGACCGCTGATTAACGAAAAATTTCTCGA 1200
Qy 1201 CGGTACTATGAAACCCCTGATGAGTTGAGTCTTTCGGAAGCATGTTACAACTG 1260
Db 1201 CGGTACTATGAAACCCCTGATGAGTTGAGTCTTTCGGAAGCATGTTACAACTG 1260
Qy 1261 ACACACAGAGATATGGACCAAAAGTGGCTTACCTGGGACCAAGTGCCTCAGAAAGAC 1320
Db 1261 ACACACAGAGATATGGACCAAAAGTGGCTTACCTGGGACCAAGTGCCTCAGAAAGAC 1320
Qy 1321 CTCATCTGGCAGACCCCTATACAGATGTAAGCCATCTCTTGTAGAGCAAAACGATTT 1380
Db 1321 CTCATCTGGCAGACCCCTATACAGATGTAAGCCATCTCTTGTAGAGCAAAACGATTT 1380
Qy 1381 GAAGCCCTAAAAGCAAAATCTCGAATCGGAGTACGGTAAAGCAGCTGGTAAGCAGC 1440
Db 1381 GAAGCCCTAAAAGCAAAATCTCGAATCGGAGTACGGTAAAGCAGCTGGTAAGCAGC 1440
Qy 1441 GCATGGCTTCTGCTACTTTTGAAGAACTCTGACAAAGCGCGGTCGCAACGGTGCA 1500
Db 1441 GCATGGCTTCTGCTACTTTTGAAGAACTCTGACAAAGCGCGGTCGCAACGGTGCA 1500
Qy 1501 CGTATACGACTGGCCCCACAAAAGACTGGGAAGTAAACACCTCAGCAACTTGGCAGG 1560
Db 1501 CGTATACGACTGGCCCCACAAAAGACTGGGAAGTAAACACCTCAGCAACTTGGCAGG 1560
Qy 1561 GTACTCAAAACACTAGAAAGTATCCAGGAGACTTTAACAGGCGCAATCAGATAACAAA 1620
Db 1561 GTACTCAAAACACTAGAAAGTATCCAGGAGACTTTAACAGGCGCAATCAGATAACAAA 1620
Qy 1621 GCATATCGTTGGCCGACCTGATTTGCTGCGCGGCTGCGGGGTGTAGAAAAAGCTGCA 1680
Db 1621 GCATATCGTTGGCCGACCTGATTTGCTGCGCGGCTGCGGGGTGTAGAAAAAGCTGCA 1680
Qy 1681 AAAGATGCTGGCCATGAGGTGCGCTTTTCAACCCGGGACGAGCGGATGCCACCGCT 1740
Db 1681 AAAGATGCTGGCCATGAGGTGCGCTTTTCAACCCGGGACGAGCGGATGCCACCGCT 1740
Qy 1741 GAGCAAAACCGATGGAAGCTTTTCGAAGCACTAGACGCGAGCGGCTGACGGCTTTAGAAC 1800
Db 1741 GAGCAAAACCGATGGAAGCTTTTCGAAGCACTAGACGCGAGCGGCTGACGGCTTTAGAAC 1800

QY 1801 TACATTAAACGGAGCATAAAGTATCCGCTGAGGAAATGCTGTAGACCGGGCGGAGCTT 1860
 Db 1801 TACATTAAACGGAGCATAAAGTATCCGCTGAGGAAATGCTGTAGACCGGGCGGAGCTT 1860
 QY 1861 CTGTGCTTTGGGACCAAGAAATGACTGCTTTGGTAGGCGGTATGCGTGTACTGGGCACC 1920
 Db 1861 CTGTGCTTTGGGACCAAGAAATGACTGCTTTGGTAGGCGGTATGCGTGTACTGGGCACC 1920
 QY 1921 AACTACGAGGTTGCGACATGAGGTGTTTCAAAATAAGCGGGTFCAGCTATFCCAATGAC 1980
 Db 1921 AACTACGAGGTTGCGACATGAGGTGTTTCAAAATAAGCGGGTFCAGCTATFCCAATGAC 1980
 QY 1981 TCTTTTGAACCTGCTAGACTCAACACTAAATGGCGGCGCAGCGATGAATCAGACAAA 2040
 Db 1981 TCTTTTGAACCTGCTAGACTCAACACTAAATGGCGGCGCAGCGATGAATCAGACAAA 2040
 QY 2041 GTTTTGAAGCAGAGACTTCAAAACTGCGAAGTAAAGTGGAGTGGCACCCTGGGTAGAC 2100
 Db 2041 GTTTTGAAGCAGAGACTTCAAAACTGCGAAGTAAAGTGGAGTGGCACCCTGGGTAGAC 2100
 QY 2101 CTGATCTTCGATCCAAATTCGAGCTAAGAGCCCTCGCAGAGTGTACGGTGTCCAGAT 2160
 Db 2101 CTGATCTTCGATCCAAATTCGAGCTAAGAGCCCTCGCAGAGTGTACGGTGTCCAGAT 2160
 QY 2161 TCTGAAGAAAGTGTGTTTAAAGATTTTGTGAAGGCTGCGCCAAAGTAAATGGACCTGGAC 2220
 Db 2161 TCTGAAGAAAGTGTGTTTAAAGATTTTGTGAAGGCTGCGCCAAAGTAAATGGACCTGGAC 2220
 QY 2221 CGTTTGTATCTGAATAA 2238
 Db 2221 CGTTTGTATCTGAATAA 2238

RESULT 2

ABS53945
 ID ABS53945 standard; DNA; 2238 BP.

XX ABS53945;

AC ABS53945;

DT 22-NOV-2002 (first entry)

DE DNA encoding Microscilla catalase protein 53CAL.

XX Gene; ss; marine bacteria; catalase; hydrogen peroxide; enzyme.

XX Microscilla furvescens.

XX Key Location/Qualifiers

FT CDS 1..2238

FT /*tag- a

FT /product- *Catalase protein*

XX US2002102680-A1.

PN 01-AUG-2002.

PD 19-JUN-2001; 2001US-0884889.

XX 03-JUL-1996; 96US-0674887.

PR 16-OCT-1997; 97US-0951844.

PR 05-OCT-1999; 99US-0412347.

XX (ROBE/) ROBERTSON D E.

PA (SANT/) SANYAL I.

PA (ADHI/) ADHIKARI R.

XX Robertson DE, Sanyal I, Adhikari R;

PI WPI; 2002-690613/74.

DR P-PSDB; ABG33041.

XX New purified catalase polypeptide useful for catalyzing the breakdown

PT of hydrogen peroxide, for modifying small molecules, and for generating

PT antibodies which bind to the polypeptide -
 XX Claim 1; Fig 6; 44pp; English.
 XX This invention relates to the DNA and protein sequences of a novel
 CC purified catalase polypeptide. The protein of the invention or its
 CC homologue is useful for catalyzing the breakdown of hydrogen peroxide.
 CC The protein sequence of the invention is useful for modifying small
 CC molecules, by mixing the protein or its fragments with a small molecule
 CC to produce a modified small molecule. The protein of the invention is
 CC also useful for catalyzing biochemical reactions, for hydrolysing
 CC glycosidic linkages and for generating antibodies which bind
 CC specifically to the protein. The nucleic acid sequences of the
 CC invention is useful as a probe to determine whether a biological sample,
 CC such as a soil sample, contains an organism having the nucleic acid or
 CC an organism from which the DNA was obtained, or for identifying an
 CC isolating related nucleic acids. An antibody that binds the cellulase
 CC protein of the invention is useful in immunoaffinity chromatography
 CC procedures to isolate or purify the protein or for detection of protein
 CC expression in a biological sample. The cellulase protein of the
 CC invention is heat stable, is heat resistant, and is able to
 CC renature and regain activity after exposure to temperatures of from
 CC about 60 to 105 degrees. The present sequence represents the DNA
 CC sequence encoding the Microscilla furvescens catalase protein of the
 CC invention.
 XX Sequence 2238 BP; 634 A; 545 C; 605 G; 454 T; 0 other;

Query Match 100.0%; Score 2238; DB 24; Length 2238;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGGAATAATCAACAACTCAGGATCTTCTAGCTATACACAAACACTGGCGGAAATGC 60
 Db 1 ATGGAATAATCAACAACTCAGGATCTTCTAGCTATACACAAACACTGGCGGAAATGC 60
 QY 61 CCTTTTACCGGAGGTTCCGTTAAGCAAGTCAGGTGGCGGACCAAAACAGGATGG 120
 Db 61 CCTTTTACCGGAGGTTCCGTTAAGCAAGTCAGGTGGCGGACCAAAACAGGATGG 120
 QY 121 TGGCCCAACATGCTCAACCTCGGCATCTTACGCCAACATTCATCGTATCGGACCCAAAC 180
 Db 121 TGGCCCAACATGCTCAACCTCGGCATCTTACGCCAACATTCATCGTATCGGACCCAAAC 180
 QY 181 GACCCGGATTTTGACTATGCCGAAGAGTTTAAAGAGCTAGATCTGCGAGCGTTAAAAAG 240
 Db 181 GACCCGGATTTTGACTATGCCGAAGAGTTTAAAGAGCTAGATCTGCGAGCGTTAAAAAG 240
 QY 241 GACCTGGCAGCGCTTAATGACAGATTCACAGGACTGGTGGCCAGCAGATTACGGTCAATAT 300
 Db 241 GACCTGGCAGCGCTTAATGACAGATTCACAGGACTGGTGGCCAGCAGATTACGGTCAATAT 300
 QY 301 GGCCCTCTCTTTATACGCATGGCGTGGACACGCGGACCTACCGTATCGGTATGGCG 360
 Db 301 GGCCCTCTCTTTATACGCATGGCGTGGACACGCGGACCTACCGTATCGGTATGGCG 360
 QY 361 CGTGTGGCGGTGGCTCCGCTCACAGCGCTTCGGCCCTCTCAATAGCTGGCCAGACAAT 420
 Db 361 CGTGTGGCGGTGGCTCCGCTCACAGCGCTTCGGCCCTCTCAATAGCTGGCCAGACAAT 420
 QY 421 GCCAATCTGGATAAAGCAGCGTTGCTTTTGGCCCATCAAAACAAAATACGGTCGAAA 480
 Db 421 GCCAATCTGGATAAAGCAGCGTTGCTTTTGGCCCATCAAAACAAAATACGGTCGAAA 480
 QY 481 ATCTCTGGCGGATCTAATGATCTACAGGAACGCTAGCTCTGGAACATATGGCTTT 540
 Db 481 ATCTCTGGCGGATCTAATGATCTACAGGAACGCTAGCTCTGGAACATATGGCTTT 540
 QY 541 AAAACTTTTGGTTTTCAGGTGGCAGCAGATGTATGGCAGCTCTGAAGAAGATGTATAC 600
 Db 541 AAAACTTTTGGTTTTCAGGTGGCAGCAGATGTATGGCAGCTCTGAAGAAGATGTATAC 600
 QY 601 TGGGGAGCAGAAACCCGAATGGCTGGGAGACAAGCGCTATGAAGGTGACCGAGAGCTCGAA 660

Db	601	 TGGGAGCAGAAACCGAATGCTGGGAGACAAGCGCTATGAAGTGAACCGAGAGCTCGAA	660
Qy	661	AATCCCTTGGGAGCCGTACAAATGGGACTCATCTATGTAAACCCCGGAAGACCCACGGC	720
Db	661	AATCCCTTGGGAGCCGTACAAATGGGACTCATCTATGTAAACCCCGGAAGACCCACGGC	720
Qy	721	AAGCCAGACCCATTCGTGTGTGGCGGTGATATTCGTGAGACTTTTGGCCGAATGGCAATG	780
Db	721	AAGCCAGACCCATTCGTGTGTGGCGGTGATATTCGTGAGACTTTTGGCCGAATGGCAATG	780
Qy	781	AATCACCAGAAACCGTGGCTCTCATAGCCGGTGGACACACCTTCGGAACCCATGGT	840
Db	781	AATCACCAGAAACCGTGGCTCTCATAGCCGGTGGACACACCTTCGGAACCCATGGT	840
Qy	841	GCTGCCGATCGGAGAAATATGTGGGCCGAGAGCGCTCCGCCGCGAGGTATTGAAGAAATG	900
Db	841	GCTGCCGATCGGAGAAATATGTGGGCCGAGAGCGCTCCGCCGCGAGGTATTGAAGAAATG	900
Qy	901	AGCCTGGGTGGAAAAACACTAGGCCACCGGACACGTCGCGATACCATCACCAGTGGGA	960
Db	901	AGCCTGGGTGGAAAAACACTAGGCCACCGGACACGTCGCGATACCATCACCAGTGGGA	960
Qy	961	CTAGAACGGCCCTGGACCAAGACCCCTACTCAATGGAGCAATTAACCTTTTGAACCTC	1020
Db	961	CTAGAACGGCCCTGGACCAAGACCCCTACTCAATGGAGCAATTAACCTTTTGAACCTC	1020
Qy	1021	TTTGGTTACAGTGGGAGCTTACCAAAAGTCCAGCTGGAGCTTATCAGTGGAAACCAAAA	1080
Db	1021	TTTGGTTACAGTGGGAGCTTACCAAAAGTCCAGCTGGAGCTTATCAGTGGAAACCAAAA	1080
Qy	1081	GACGGTCCGGGGTGGCACCATACCGGATGCACATGCCAGCAAGTCGCACGCTCCA	1140
Db	1081	GACGGTCCGGGGTGGCACCATACCGGATGCACATGCCAGCAAGTCGCACGCTCCA	1140
Qy	1141	TTTATGCTCACTACGGACCTTGGCGCTCGCATGGACCCCTGATTACGAAAAAATTTCTCGA	1200
Db	1141	TTTATGCTCACTACGGACCTTGGCGCTCGCATGGACCCCTGATTACGAAAAAATTTCTCGA	1200
Qy	1201	CGGTACTTATGAAAACCCCTGATGAGTTTGCAGATGCTTTCGCGAAAGCATGGTACAAACTG	1260
Db	1201	CGGTACTTATGAAAACCCCTGATGAGTTTGCAGATGCTTTCGCGAAAGCATGGTACAAACTG	1260
Qy	1261	ACACACAGAGATATGGGACCAAAAGTGGCTACTCTGGGACCGAGTGCCTCAGGAAGAC	1320
Db	1261	ACACACAGAGATATGGGACCAAAAGTGGCTACTCTGGGACCGAGTGCCTCAGGAAGAC	1320
Qy	1321	CTCATCTGGCAAGACCCATACCAGATGTAAAGCCATCCCTTGTAGACGAAAAACGATATT	1380
Db	1321	CTCATCTGGCAAGACCCATACCAGATGTAAAGCCATCCCTTGTAGACGAAAAACGATATT	1380
Qy	1381	GAAGGCCCTAAAGCCAAAATCCTGGAAATCGGGACTGACGGTAAAGCGAGCTGGTAAAGCAGC	1440
Db	1381	GAAGGCCCTAAAGCCAAAATCCTGGAAATCGGGACTGACGGTAAAGCGAGCTGGTAAAGCAGC	1440
Qy	1441	GCATGGGCTTTCATCTACTTTTAGAAAACCTCTGACAAGCGCGCGGTGCCAACGGTGGCA	1500
Db	1441	GCATGGGCTTTCATCTACTTTTAGAAAACCTCTGACAAGCGCGCGGTGCCAACGGTGGCA	1500
Qy	1501	CGTATACGACTGGCCCGCACAAAAGACTGGGAAGTAAACAACCCCTCAGCAACTTGGCAGG	1560
Db	1501	CGTATACGACTGGCCCGCACAAAAGACTGGGAAGTAAACAACCCCTCAGCAACTTGGCAGG	1560
Qy	1561	GTACTCAAAACACTAGAAAGTATCCAGGAGACTTTAAACCGACGCGCAATCAGATACAAA	1620
Db	1561	GTACTCAAAACACTAGAAAGTATCCAGGAGACTTTAAACCGACGCGCAATCAGATACAAA	1620
Qy	1621	GCAGTATCGTTGCCCGCACTGATTGTCTGGCCGGCTGTCCGGGTGTAGAAAAAGCTGCA	1680
Db	1621	GCAGTATCGTTGCCCGCACTGATTGTCTGGCCGGCTGTCCGGGTGTAGAAAAAGCTGCA	1680
Qy	1681	AAAGATGCTGGCCATGAGGTGCAGTGCCTTTCAACCCGGGACGACGGATGCCACCGCT	1740

Db	1681	AAAGATGCTGGCCATGATGAGGTGCAGGTGCCCTTTCAACCCGGGACGAGCGGATGCCACCGCT	1741
Qy	1741	GAGCAAAACCGATGTGGGAAGCTTTTCGAAGCACTAGAGCCAGCGCTGACGGCTTTAGAAAC	1800
Db	1741	GAGCAAAACCGATGTGGGAAGCTTTTCGAAGCACTAGAGCCAGCGCTGACGGCTTTAGAAAC	1800
Qy	1801	TACATTAACCCGGAGCAATAAGTATCCGCTGAGGAATGCTCGTAGACCCGGCGCAGCTT	1860
Db	1801	TACATTAACCCGGAGCAATAAGTATCCGCTGAGGAATGCTCGTAGACCCGGCGCAGCTT	1860
Qy	1861	CTGTGCGTTTCGGCACCAGAAATGACTGCTTTGGTAGGCGGTATGCGTACTTGGGCACC	1920
Db	1861	CTGTGCGTTTCGGCACCAGAAATGACTGCTTTGGTAGGCGGTATGCGTACTTGGGCACC	1920
Qy	1921	AACATCACACGGTTCGGCAGCATGAGTGTGTTTACAAATAAGCCGGGTACGCTATCCAATGAC	1980
Db	1921	AACATCACACGGTTCGGCAGCATGAGTGTGTTTACAAATAAGCCGGGTACGCTATCCAATGAC	1980
Qy	1981	TTCTTTTGAACCTGTCTAGACCTCAACACTAAATGGCGAGCCAGCGATGAATCAGACAAA	2040
Db	1981	TTCTTTTGAACCTGTCTAGACCTCAACACTAAATGGCGAGCCAGCGATGAATCAGACAAA	2040
Qy	2041	GTTTTTCGAAGCAGAGACTTCAAACTGGCGAAGTAAAGTGGAGTGGCAGCCCGGGTAGAC	2100
Db	2041	GTTTTTCGAAGCAGAGACTTCAAACTGGCGAAGTAAAGTGGAGTGGCAGCCCGGGTAGAC	2100
Qy	2101	CTGATCTTCGGATCCCAATCCGAGCTTAAGAGCCCTCGCAGAACTGTACGGCTGTGCAGAT	2160
Db	2101	CTGATCTTCGGATCCCAATCCGAGCTTAAGAGCCCTCGCAGAACTGTACGGCTGTGCAGAT	2160
Qy	2161	TCTGAAGAAAAGTTTGTAAAGATTTTGTGAAGCGCTTGGGCCAAAGTAATGGACCTGGAC	2220
Db	2161	TCTGAAGAAAAGTTTGTAAAGATTTTGTGAAGCGCTTGGGCCAAAGTAATGGACCTGGAC	2220
Qy	2221	CGGTTTGATCTGAATAA 2238	
Db	2221	CGGTTTGATCTGAATAA 2238	
RESULT 3			
ID	AAT90400 standard; DNA; 2331 BP.		
XX	AAT90400;		
XX	25-MAR-2003 (updated)		
DT	21-JAN-1998 (first entry)		
XX	Mycobacterium tuberculosis wild-type katG gene.		
DE	Mycobacterium tuberculosis.		
KW	Isoniazid resistance; restriction fragment length polymorphism;		
KW	RFLP; katG gene; catalase; peroxidase; multiple drug resistance;		
KW	isonicotinic acid hydrazide; INH; tuberculosis; antibiotic;		
XX	tuberculostatic; ss.		
OS	Mycobacterium tuberculosis.		
Key	-	Location/Qualifiers	
FT	CDS	70..2292	
FT	mutation	/*tag= a	
FT		859	
FT		/*tag= b	
FT		/note= "G to C transversion at base 859 alters	
FT		codon 284 from GCG (Ala) to ACG (Thr)	
FT		leading to isoniazid resistance, and	
FT		eliminates a CfoI restriction site"	
FT	mutation	1013	
FT		/*tag= C	
FT		/note= "G to C transversion at base 1013 alters	
FT		codon 315 from AGC (Ser) to ACC (Thr)	
FT		leading to isoniazid resistance, and	
FT		produces a new MspI restriction site"	
FT	mutation	1079	

QY 1410 GGGACTACGGTAAAGGAGCTGTTAAGCAGCGGATGGGCTTCTGCACTACTTTTAGAAA 1469
 DB 1464 GGGATGACTGTCTACAGACTAGTTTCGACCGATGGGCGCGCTGCTGCTCCGTGG 1523
 QY 1470 CTCTGCAAGCGGGGCTGCCAAGCGTGCACGTATACAGCTGSCGCCCAAAAAGACTG 1529
 DB 1524 TAGCGACAGCGGGCGGCCAAGCGTGGTGGATCCGCTCGAGCCACAAGTCGGGGT 1583
 QY 1530 GGAAGTAAACAACCCCTCA---GCAACTTGGCCAGGGTACTCAAAACACTAGAAGGTATCCA 1586
 DB 1584 GGAGGTCAACGACCCCGACGGGATCTGCGCAAGGTCAATTCGCACCTCGAAGAGATCCA 1643
 QY 1587 GGAGGACTTTAACGAGCGGAATACAGATACAAAGCAGATATCGTTGGCCGACCTGATGT 1646
 DB 1644 GGAGTCAATTCAACTCCGCGCGCGGGAACATCAAAAGTCTGCTTCCGCCACCTCGTCT 1703
 QY 1647 GCTGGCGGCTGTGCGGCTGTAGAAAAGCTGCAAAAGATGCTGGCCATCAGTGCAGGT 1706
 DB 1704 GCTCGGTGGCTGTGCGGCCATAGAGAAAGCAGCAAGCGCGCTGGCCACCAACATCAGGT 1763
 QY 1707 GCCTTTCAACCCGGGAGCGGATGCCACCGCTGAGCAAAACGATGTGGAAGCTTTTCA 1766
 DB 1764 GCCTTCAACCCGGGCGCGGATGCTGCGAAGCGAAGCTGCTTACGCTCAGTGCCCTGAGATGAC 1943
 QY 1887 TGTCTTGGTAGCGGTATGCTGTACTGCGCACCACCTAGCAGCGTTTCGACGATGGAGT 1946
 DB 1944 GGTGCTGTAGTGGCTGCGGCTCTCGCGCAAACTACAAGCGCTTACCGCTGGCGCT 2003
 QY 1947 GTTTACAAATAAGCGGCTCAGCTATCAATGACTTTTGTAAACCTGTAGACCTCAA 2006
 DB 2004 GTTACCGAGGCTCCGAGTCACTGACCAACGACTTCTTGTGAACCTGTGTCGACATGGG 2063
 QY 2007 CACTAATGCGAGCGGAGGATGAATCAGACAAAGTTTGTGAAGGAGAGACATTCAAAC 2066
 DB 2064 TATCACTGGAGCCCTCGCAGCAGATGACGGGACCTACCGGCAAGATGCG---AG 2120
 QY 2067 TGGCGAAGTAAAGTGGAGTGGCAGCGGATGACCTGATGATTCGGATCCAAATCCGAGCT 2126
 DB 2121 TGGCAAGTGAAGTGGAGCGGCGCGCTGGACCTGCTTCCGGTCCCAACTCGGAGTT 2180
 QY 2127 AAGAGCCTCGCAGAGTGTACGCTGTGCGATCTCTGAAGAAAGTTTGTAAAGATTT 2186
 DB 2181 GCGGCGCTGTGCGAGTGTATGGCGCGATGACGCGGACCGCAAGTTCGTGCGAGACTT 2240
 QY 2187 TGTGAAGGCTGGGCGCAAGTAATGACCTGGACCGGTTTGTATCTG 2232
 DB 2241 GCTGCTGCTGGACAGGTGATGAACCTCGACAGGTTGACGTG 2286

RESULT 4

AAX00817

ID AAX00817 standard; DNA; 2331 BP.

XX AAX00817;

AC AAX00817;

XX 26-MAR-1999 (first entry)

DT M. tuberculosis catalase peroxide (katG) gene sequence.

XX M. tuberculosis catalase peroxide (katG) gene sequence.

XX Catalase-peroxide; katG; isonicotinic acid hydrazide; tuberculosis;

KW INH; diagnosis; detection; ss.

XX Mycobacterium tuberculosis.

OS Mycobacterium tuberculosis.

XX Key

XX Location/Qualifiers

XX

XX

XX

XX

XX

XX

XX

XX

FT CDS 70..2292
 FT /*tag= a
 FT /gene= "katG"
 FT /product= "catalase peroxide"
 FT /note= "the start codon is not indicated"
 XX
 PN W09850585-A1.
 XX
 XX 12-NOV-1998.
 XX
 XX 06-MAY-1998; 98WO-US09285.
 XX
 XX 07-MAY-1997; 97US-0852219.
 XX
 XX (MAYO-) MAYO FOUNDATION.
 XX
 XX Cockerill FR, Kline BC, Uhl JR;
 XX
 XX WPI; 1999-070099/06.
 DR P-PSDB; AAW95398.
 XX
 XX Detection of Mycobacterium tuberculosis - by amplifying katG gene
 XX and detecting specific fragment, and optionally identifying
 XX INH-resistant strains by detecting specific mutation
 XX
 XX Claim 3; Fig 7; 83pp; English.
 XX
 XX The invention relates to a novel method of detecting Mycobacterium
 XX tuberculosis. The method comprises amplifying the DNA in the samples to
 XX generate a detectable amount of amplified DNA comprising a catalase-
 XX peroxide (katG) DNA fragment with sequence of bases 904-1533 of the
 XX M. tuberculosis strain H37Rv MC katG gene sequence. The method optionally
 XX further comprises determining if the katG DNA fragment has a serine to
 XX threonine mutation in codon 315 (S315T mutation), indicative of an
 XX isonicotinic acid hydrazide (INH)-resistant M. tuberculosis strain. The
 XX method can be used to detect M. tuberculosis in biological fluids,
 XX especially human sputum, useful to diagnose tuberculosis. This disease is
 XX a major cause of human morbidity and mortality, and conclusive diagnosis
 XX and subsequent treatment depends on identification of the etiologic agent
 XX M. tuberculosis. INH has been used in tuberculosis treatment, but
 XX INH-resistant strains have emerged; the method allows such drug-resistant
 XX strains to be identified. The present sequence represents the wild-type
 XX M. tuberculosis katG gene sequence.
 XX
 SQ Sequence 2331 BP; 455 A; 734 C; 763 G; 379 T; 0 other;
 Query Match 37.7%; Score 844.4; DB 20; Length 2331;
 Best Local Similarity 63.0%; Pred. No. 1.8e-253;
 Matches 1390; Conservative 0; Mismatches 801; Indels 15; Gaps 5;
 QY 30 TAGGTATAACACAAACACTGGCGAAATGCCCTTTTACCGGAGTTTCGCTTAAGCAAAG 89
 DB 93 TACAGAAACACCAACCGGAGCGCTAGCAACGGCTGTCCGCTCATATGAATA 152
 QY 90 TGCAGTGGCGGACCAACAAACAGGATTTGGTGGCCCAACATGCTCAACCTCGGCACTTT 149
 DB 153 CCCGCTCGAGGGCGGGAACCAAGGACTGGTGGCCCAACCGGCTCAATCTGAAGGTACT 212
 QY 150 ACGCCAACTTCATCGCTATCGGACCAACAGACCGGATTTTACTATATGCCAAGATT 209
 DB 213 GCACAAACCCCGGCGTCTGCTGACCCGATGGGTGGGCGCTTCGCTATATGCCGAGGT 272
 QY 210 TAAGAAGCTAGATCTGGCAGCGGTTAAAGAGACCTGGCAGCGCTAATGACAGATTACCA 269
 DB 273 CGCGACCATCGAGTTGACGCCCTGACCGCGGACATCGAGAAAGTATGACCACTCGCA 332
 QY 270 GGACTGGTGGCCAGCAGATTACGCTATTATGCCCTTCTTTATACGATGGCGTGCCA 329
 DB 333 GCGCTGGTGGCGCGGACTACGCCACTACGGCCGCTCTTTATCCGATGGCTGGCA 392
 QY 330 CAGCGCGGCGACCTACCGTATCGGTATGCGGCTGGTGGCGGCTGCCGCTCACAGCG 389
 DB 393 CGTGGCGGCACTTACCGCATCCAGACGCGCGCGCGCGCGCGGCGGCGGCGCATGCAAGC 452

QY 390 CTTTCGGCGCTCTCAATAGCTGGCCAGACAAATGCCAATCTGGATAAAGACACGCTTGTCTTCT 449
 Db 453 GTTCGGCGCGCTTAACAGCTGGCCGACAAACGACGCTTGGCAAGCGCGCGCGCTGCT 512
 QY 450 TTGGCCCATCAACAAATAACGCTCGAATAATCTCTCGGGGGATCTAATGATCTAC 509
 Db 513 GTGGCCGGTCAAGAAGTACGGCAAGAGCTCTCATGGGGGACCTGATTTGTTTCGC 572
 QY 510 AGGAACGCTAGCTCTGGAACATATGGCTTTTAAACCTTTTGGTTTTGACAGTGGCAGAGC 569
 Db 573 CGGCAACTGCGCGCTGAATCGATGGCTTCAAGACGTTTCGGGTTTCGGCTTCGGCGGGT 632
 QY 570 AGATGATGGGACCTTGAAAGATGATATCTAGTGGGAGCAGAAACGATGGCTGGGAGA 629
 Db 633 CGACCACTGGGAGCC---CGATGAGTCTATTGGGGCAAGGAAGCACCTGGCTCGGGCA 689
 QY 630 CAAGCGCTATGAAGCTGACCGAGCTCGAATAATCCCTGGGAGCGGTACAATGGCACT 689
 Db 690 TGAGCGTTACAGCGGTAAAGGGATCTGGAGAACCGCTGGCCGCGTGCAGATGGGGCT 749
 QY 690 CATCTATGAACCCCGAAGGACCCAAACGCGCAAGCAGACCCCTATCGCTGCTGCGGTGA 749
 Db 750 GATCTACGTGAACCCCGAGGGCGGAACGCGCAACCGGACCCCATGGCCGCGCGCTCGA 809
 QY 750 TATTCGTGAGACTTTTGGCGGAATGCAATGCAATGAGAGAAACCGCTGCTCTCATAGC 809
 Db 810 CATTCGCGAGAGCTTTTCGCGCATGGCCATGACACAGCTCGAACAACGCGCGCTGATCGT 869
 QY 810 GGTGTGACACACCTTCGGAATAACCCATGCTGCTGCGGATGCGGAGAAATATGTGGCGG 869
 Db 870 CGCGCTACACTTCGCTAAGACCCATGCGG---CGGCGCGCGGCTGCTGGTCGCGCC 926
 QY 870 AGAGCTGCGCGCGAGGTATTGAAGAAATGAGCCCTGGGTGGGTGGAATAACACCTTACGCGAC 929
 Db 927 CGAACCAGGCTGCTCCGCTGGAGCAGATGGCTTGGGCTGGAACAGCTCGTATGCGAC 986
 QY 930 CGGACAGCTGGGATACCATCAGCTGAGCTAGAGGCGCTGAGCAACAGACCCCTTAC 989
 Db 987 CGGAACCGGTAGGAGCGCATACCAAGCGGCTCGAGGTGCTATGACGCAACACCCCGAC 1046
 QY 990 TCAATGAGCACTAATCTTTTCAAAACCTCTTTGGTTAGCTGGGAGCTTACCAAAAG 1049
 Db 1047 GAATGGGACACAGTTTCTCGAGATCTGTAGCGCTACGAGTGGAGCTGACGAAGAG 1106
 QY 1050 TCAGCTGAGCTTACGTGGAACCAAAAGACGCTGCGGGCTGGCACCACTACCGGA 1109
 Db 1107 CCCTGCTGGCGCTTGGCAATACACCGCAAGGACGCGCGGTGCGCGGACCACTCCCGGA 1166
 QY 1110 TGCACATGATCCAGCAAGTCCGACCGCTCCATTTATGCTCACTACGAGCTGCGCTGGG 1169
 Db 1167 CCGCTTTCG---CGGGGCCAGGGCGCTCCCGGACGATGCTGGCCACTGACCTCTCGCTGG 1223
 QY 1170 CATGGACCTGATTACGCAAAATTTCTCGACGGTACTATGAAACCCCTGATCAGTTTCG 1229
 Db 1224 GTTGGATCCGATCTATGAGCGGATACGCGTCTGCTGGCTGGAACCCCGGAGAAATGGC 1283
 QY 1230 AGATGCTTTGGCAAGAGCTGTACAACTGACACAGAGATATGGGACCAAGGTGGC 1289
 Db 1284 CGAGAGTTGGCAAGCGCTGTACAGCTGTACCGAGCATGGGTCCCGTTGGGAG 1343
 QY 1290 CTACCTGGGACAGAGTGGCTCAGGAAGACCTCATCTGCGCAAGACCCCTATACCATGAT 1349
 Db 1344 ATACCTTGGGCGCTGTGCTCCCAAGCAGACCCCTGCTGTGCGAGGATCCCGTCCCTGGGT 1403
 QY 1350 AAGCCATCTCTTGTAGACGAACCAATATTGAAGGCTTAAAGCCAAATCTCGAATC 1409
 Db 1404 CAGCCAGGACTGCTGGGGAAGCGGAGATTGCGAGCTTAAAGCGCAGATCCGGGCATC 1463
 QY 1410 GGGACTGACGGTAAGCGAGCTGTAAAGCAGCGCATGGGTCTCTGCTACTCTTTTGAATA 1469
 Db 1464 GGGATTGACTGCTCAGAGCTAGTTTCGACCGCATGGGCGCGCGCTGCTGCTCGGTGG 1523

QY 1470 CTCTGACAAAGCGCGCGGTGCCAACCGGTGCACGCTATACGACTGGCCCCACAAAAAGACTG 1529
 Db 1524 TAGCGCAAGCGCGCGGCCCAACGCTGTGTGCTATCCGCTCGACGCCAAGTCGGGTG 1583
 QY 1530 GGAAGTAAACAACCCCTCA---GCAACTTCCAGGCTACTCAAAACACTAGAGGTATCCA 1586
 Db 1584 GGAGGTCAACAGACCCCGACGGGATCTGGCAAGGTCATTCGCACCCTCGAAGATCCA 1643
 QY 1587 GGAGGACTTTAAACAGCGCAATCAGATAACAAAGCAGATATGTTGGCGGACCTGATTTGT 1646
 Db 1644 GGAGTCATTCAACTCCGCGCGCGCGGGAACATCAAAAGTGTCTTTCGCGGACCTCGCTGT 1703
 QY 1647 GCTGGCGCGCTGTGGGGTGTAGAAAAGCTGCAAAAAGATGCTGGCCATGAGGTGAGGT 1706
 Db 1704 GCTGGGTGCTGTGGCGCATAGAAAAGCAGCAAAAGCGGCTGGCGCAACATCACTCGGT 1763
 QY 1707 GCCTTTCAACCCGGGACGAGCGATGCCACCGTGAGCAAAACGATGTGGAGCTTTTCCA 1766
 Db 1764 GCGCTTCAACCCGGGCGGACGGATGCTGCGAGGAACAAACCGAGTGGATCTTTGCG 1823
 QY 1767 AGCACTAGAGCCAGCGGCTGACGGCTTTAGAACTACATTTAAACCGGAGCATAAAGTATC 1826
 Db 1824 CGTGTGGAGCCCAAGGACAGATGGCTTCGAAACTACCTCGGAAGAGGCAACCCGTTGCC 1883
 QY 1827 CGTGAAGGAATGCTCTGTAGACCGGCGGAGCTTCTGTGCTTTTCGGCACCAAGATGAC 1886
 Db 1884 GSCGAGTACATGCTCTGACAAAGCGAACCTGCTTAGCTCAGTCCGCTGAGATGAC 1943
 QY 1887 TGTCTTGGTGGGCTGATGCTGCTACTGGGACCAACTAGAGGTTTCGACGATGAGT 1946
 Db 1944 GTGTCTGTGGTGGCTGCTGCGCGCAAACTACAAGCGTTTACCCTGGGCGT 2003
 QY 1947 GTTTCAAAATAAGCCGGGTCAGCTATCCATGACTTCTTTGTAACCTGCTAGACCTCAA 2006
 Db 2004 GTTACCGAGGCTCCGAGTCACTGACCAAGAGCTTCTTCGTGAACCTGCTCGACATGG 2063
 QY 2007 CACTAAATGGCGAGCCAGCGATCAACAGAAAGTTTTTGAAGGACAGACTTCAAAAC 2066
 Db 2064 TATCACTGGGAGCTCCGACGAGATGACGGGACCTACCGAGGCAAGGATGCG---AG 2120
 QY 2067 TGGCAAGTAAAGTGGAGTGGCACCGGCTAGACCTGATCTTCGGATCCCAATTCGAGCT 2126
 Db 2121 TGGCAAGTGAAGTGGAGCGGACCGCGGCTGGACCTGCTTCGGGTCCCACTCGGAGTT 2180
 QY 2127 AAGAGCCCTCGCAAGAGTGTACGGCTGTGCAGATCTGGAAGAAAGTTTGTAAAGATT 2186
 Db 2181 GCGGGCGCTTGTGAGGTCTATGGCGCGATGACGCGCAGCGCAAGTTCGTGCGAGACTT 2240
 QY 2187 TGTGAAGGCTTGGGCAAAAGTAATGACCTGGACCGGTTTGTGATCTG 2232
 Db 2241 CGTCTGCTGGGCAAGGTGATGAACCTCGACAGGTTTCGAGGTG 2286

RESULT 5

ABX09140/c
 ID ABX09140 standard; DNA; 82993 BP.

XX
 AC ABX09140;

XX
 DT 08-APR-2003 (first entry)

XX
 DE Mycobacterium tuberculosis H37Rv BAC clone BAC-rv221.

XX
 XX Mycobacteriosis; survival; virulence; protective antigen; vaccine;
 KW mycobacterial disease; tuberculosis; leprosy; ds; cosmid.

XX
 OS Mycobacterium tuberculosis.

XX
 PN WO200274903-A2.

XX
 PD 26-SEP-2002.

XX
 PF 22-FEB-2002; 2002WO-IB01973.

QY 810 GGTGTGACACACTTCGGAAGAAACCCATGGTGTGCTGCCGATGCGGAGAAATATGTGGCGG 869
 Db 2152569 CGCGGTCTACACTTCGCTTAAGACCCATGCGCG- --CGCGCGCGCGATCTGCTGCGCGC 2152513
 QY 870 AGAGCTCGCGCGCAGGTATTGAAGAAATGAGCCTGGGGTGGAAAAACACCTACGGCAC 929
 Db 2152512 CGAACCCGAGGTGCTCCGCTGGAGCAGATGGCTTGGCTGGAGAGTCTGTATGGCAC 2152453
 QY 930 CGGACACCTGGGGTACCATTACCATCAGTGGAGCTAGAGAGCGCTGGACCAAGACCCCTAC 989
 Db 2152452 CGGAACCGGTAAAGACCGCATCACCAGCGCATCGAGGTCTGTATGACGACCAACCCCGAC 2152393
 QY 990 TCAATGGAGCAATAACTTTTTTGAACCTCTTTTGGTTACGAGTGGAGCTTACCAAAAG 1049
 Db 2152392 GAAATGGGACACAGTTTCTCGAGATCTGTACGGCTACGAGTGGAGTGAAGAG 2152333
 QY 1050 TCCAGCTGAGCTTATCACTGGAAACCAAGACGGTGGCGGCTGGACCAATACCGA 1109
 Db 2152332 CCTGCTGGCGCTTGGCAATACACCGCAAGGACGGCGGCTGGCGGACCATCCCGA 2152273
 QY 1110 TGCACATGATCCAGCAAGTCCGACGCTCCATTATGCTCACTACGAGCTGGCGTGG 1169
 Db 2152272 CCGCTTCG- --GGGGCCAGGGCGCTCCCGACGATGCTGGCCACTGACCTCTCGCTGG 2152216
 QY 1170 CATGGACCTGATTAGCAAAAAATTTCTCGAGGTACTATGAACCCCTGATGAGTTTC 1229
 Db 2152215 GGTGATCCGATCTATGAGCGATCAGCGCTCGCTGGCTGGAAACCCCGAGGAATTGGC 2152156
 QY 1230 AGATGCTTTTCGGAAGCATGTGTAACAACTGACACACAGATATGGGACCAAGTGG 1289
 Db 2152155 CGACGATTTGCGCAAGCGCTGTACAACTGATCCACGAGACATGGGTCGCTGGAG 2152096
 QY 1290 CTACCTGGGACCAAGATCCCTCAGGAAGACCTCTCTGGCAAGACCCCTATACGAGAT 1349
 Db 2152095 ATACCTTGGCGCTGTGCTCCCAAGCAGACCTGCTGTGSCAGGATCCGCTCGCT 2152036
 QY 1350 AGCCATCTCTGTAGACGAAACGATATTGAAGCCCTAAAGCCCAAAATCTGGAATC 1409
 Db 2152035 CAGCCACGACCTCTGTCGGCGAGCCGAGATTGCCAGCTTAAGAGCCAGATCGGGCATC 2151976
 QY 1410 GGGACTGACGTTAAGCGACTGGTAAGCACGCTGCTGCAAGCTTCTGCACTACTTTAGAA 1469
 Db 2151975 GGGATTGACTGTCTACACCTAGTTTCGACCGCATGGGCGGCGCTGCTGCTCGTGG 2151916
 QY 1470 CTCTGACAAAGCGCGGTGCCAAGCTGACATATACGACTGGCCGACCAAAAGACTG 1529
 Db 2151915 TAGCGAAGACGCGCGCGCGCAACGGTGTGCTCGATCCGCTGACGCCCAAGTGGGTG 2151856
 QY 1530 GGAATTAACACCCCTCA- --GCAACTTCCAGGGTACTCAAAACACTAGAGGATATCA 1586
 Db 2151855 GAGGTCAACGACCGCGCGGATCTGCGCAAGGTCTTCGACCCCTGGAGAGATCCA 2151796
 QY 1587 GGAGGACTTTAACGAGCGCAATCAGATACAAAGCAGTATCGTTGGCGGACCTGATTTC 1646
 Db 2151795 GGAGTCACTCACTCGCGCGCGCGGGGAACATCAAGTGTCTTGGCGGACCTGCTGT 2151736
 QY 1647 GCTGCGCGCTGTGCGGTGTGAAAGAGCTGCAAAAGATGCTGGCCATGAGTGGCAGT 1706
 Db 2151735 GCTGCGTGTGCTGCGCATAGAGAAAGCAGCAAGGCGCTGGCCACACATCAGCT 2151676
 QY 1707 GCTTTCAACCGGACGAGCGGATGCCACCGCTGAGCAACCGATGTGGAGCTTTCA 1766
 Db 2151675 GCGCTTACCGCGCGCGCGGATGCGTGGAGGAACAAACCGAGTGGAAATCTCTTC 2151616
 QY 1767 AGCACTAGACGCGGCTGACGGCTTTAGAACTACATTAACCGGACCATAAAGTATC 1826
 Db 2151615 CTTGCTGGAGCCCAAGCGAGATGGCTTCGGAACCTACCTCGGAAGGGAACCGTGGC 2151556
 QY 1827 CGCTGAGGAATGCTCGTAGACGCGCGAGCTTCTGTGCTTTTGGCACCAAGATGAC 1886
 Db 2151555 GCGCGAGTACATGCTGCTCGACAGCGGAACCTGCTTACGCTCAGTGCCTGAGATGAC 2151496
 QY 1887 TCGTTTGGTAGCGGTATCCGTGTACTGGGCAACCACTACGACGGTTCCGACGATGGAGT 1946

Db 2151495 GGTGCTGGTAGGTGGCTTCGCGCTCCTCGCGCGCAAACTACAAGCGCTTACCGCTGGCGT 2151436
 QY 1947 GTTTACAAATAAGCCGGGTGAGTATCCATGACTTCTTTGTAAACCTGCTAGACCTCA 2006
 Db 2151435 GTTCACCGAGGCTCCGAGTCTCTGACCAACGACTTCTTCGTGAACCTGCTGACATGG 2151376
 QY 2007 CACTAAATGGCAGGCGAGCGATGAATCAGACAAAGTTTTTTGAAGGAGAGATTCAAAAC 2066
 Db 2151375 TATCACTGGAGCCCTCCGACGATGCGGGACCTACCGAGGCAAGGATGGC- --AG 2151319
 QY 2067 TGGCGAAGTAAAGTGGAGTGGCACCCGCGGTAGACCTGATCTTCGGATCAAAATCCGAGT 2126
 Db 2151318 TGGCAAGGTGAAGTGGACCGCGCGCTGGACCTGGTCTTCGGGTCCAACTCGGAGTT 2151259
 QY 2127 AAGAGCCCTCGAGAGTGTACGGCTGTGAGGATCTGAGAAAGTTTTTTAAAGATTT 2186
 Db 2151258 GCGGCGCTGTGAGAGTCTATGGCGCGCATGACGCGAGCGAAGTTCTGCGAGGACTT 2151199
 QY 2187 TGTGAAGGCTGGGCGCAAGTAAATGACCTGGACCGGTTTGATCTG 2232
 Db 2151198 CGTGGCTGCTGGGCAAGGTGATGAACCTCGACAGGTTTCGAGGTG 2151153
 RESULT 7
 AA199682/c
 ID AA199682 standard; DNA; 4411529 BP.
 XX AA199682;
 XX
 DT 15-JAN-2002 (first entry)
 XX
 DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
 XX
 KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
 KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN US6294328-B1.
 XX
 PD 25-SEP-2001.
 XX
 PF 24-JUN-1998; 98US-0103840.
 XX
 PR 24-JUN-1998; 98US-0103840.
 XX
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Fleischmann RD, White OR, Fraser CM, Venter JC;
 XX WPI; 2001-647261/74.
 DR
 XX
 PT
 PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
 PT determining the nucleotide sequence of the strain at positions in the
 PT genome corresponding to positions where M. tuberculosis strains CDC
 XX 1551 and H37Rv differ
 PS Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
 XX
 CC The invention relates to evaluating strain variation within and between
 CC different populations of the tuberculosis bacterial pathogen,
 CC Mycobacterium tuberculosis or related Mycobacterium by determining the
 CC nucleotide sequence of the first strain at positions in the complete
 CC sequence of the genome that correspond to positions that differ in the
 CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and
 CC H37Rv (AA199682). The method is useful for evaluating strain variation of
 CC M. tuberculosis and has valuable application in the fields of
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic
 CC monitoring.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from USPTO
 CC at seqdata.uspto.gov/sequence.html?DocId=6294328B1.

XX	Sequence	4411529	BP: 758565	A: 1449983	C: 1444602	G: 758379	T: 0	other:
SQL	Query Match	37.7%	Score	844.4	DB	22	Length	4411529;
	Best Local Similarity	63.0%	Pred. No.	1.6e-251;				
	Matches	1390;	Conservative	0;	Mismatches	801;	Indels	15;
								Gaps
QY	30	TACGATATACACAAACACTCGGGGAAATGCCCTTTTACGGAGGTTTCGCTTAAGCAAG	89					
DB	2156086	TACAGAAACACCAACCCGAGCGCTAGCAACGGCTGTCCCGTGTGGTTCATATGAAATA	2156027					
QY	90	TGACGGTGGCGGCACCAACACAGGATTCGTGGCCCAACATGCTCAACCTCGGCATCTT	149					
DB	2156026	CCCCTCGAGGGCGGGGAAACACGAGACTGGTGGCCCAACCGGCTCAATCTGAAGGTACT	2155967					
QY	150	ACGCAACATTCATTCGCTATTCGACCCCAACAGCCGATTTTGTACTATGCCGAAGATT	209					
DB	2155966	GCACCAAAACCCGGCGCTGCTGACCCGATGGGTGGCGGCTTGACATATGCCCGGAGGT	2155907					
QY	210	TAGAAGCTAGATCTGGCAGCGGTTAAAGAGACTTGGCAGCGCTAATACAGATTACACA	269					
DB	2155906	CGGCACCATCGAGTTGACGCCCTTGACCGCGGACATCGAGGAAGTGTATGACCACTCGCA	2155847					
QY	270	GGACTGTGGCCAGCAGATTACGGTCAATATGGCCCTCTCTTATACGATGGGTGGCA	329					
DB	2155846	CGCGTGTGGCCCGCGGACTACGGCCACTACGGGCGCTGTTATCCGATGGGTGGCA	2155787					
QY	330	CAGCGCCGACCTTACCGTATCCGGTATCCGGTGGTGGCGGTGCTCCGGCTCACAGCG	389					
DB	2155786	CGCTGCCGCGACCTACCGCATCCACGACGCGCGCGGCGCGCGGGGCGGCGATGCAGCG	2155727					
QY	390	CTTGGCGCCTCTCAATAGCTGGCCAGACATGCCAATCTGATTAAGACACGCTTGCTTCT	449					
DB	2155726	GTTTGGCGCGCTTAAACAGCTGGGCCGACACAGCGAGCTTGGCAAGCGCGCGGCTGCT	2155667					
QY	450	TTGGCCCATCAACAAATAACGCTGCAAAATCTCTCGCGCGGATCTAATGATACTCAC	509					
DB	2155666	GTGGCCCGTCAAGAAGATGACGCAAGAGCTCTCATGGCGGACCTGATGTTTTCGC	2155607					
QY	510	AGAAACAGTACTCTGAAACTATGGCTTTTAAACTTTTGGTTTTGCAGGTGGCAGAC	569					
DB	2155606	CGGCAACTGGCGCTGAATCGATGGCTTCAAGACGTTCCGGTTCCGGCTTCGGCGGGT	2155547					
QY	570	AGATGTATGGAGCCTGAAGAAGATGATATCTGGGAGCAGAAACCGAATGGCTGGAGA	629					
DB	2155546	CGACCACTGGGAGCC---CGATAGGTCTATTGGGGCAAGAACCCACCTGGCTCGGCA	2155490					
QY	630	CAAGCGCTATGAAGTGACGAGAGCTCGAAATCCCTCGGAGCCCTACAAATGGAGCT	689					
DB	2155489	TGACGGTTACAGCGGTAAAGCGGGATCTGGAGAACCCGCTGGCGCGGTGCAGATGGGCT	2155430					
QY	690	CATCTATTAACCCGAGGACCAACGCAAGCCAGACCTATCCTGCTGCGCGTGA	749					
DB	2155429	GATCTAGTGAACCGGAGGGCGGACGCAACCGGACCCCATGCGCGCGGCGTTCGA	2155370					
QY	750	TATTCGTGAGACTTTTGGCGAATGGCAATGAATGACGAAGAACCGTGGCTCTCATAGC	809					
DB	2155369	CATTCGCGAGAGCTTTCCGCGCATGCGCATGTAAGACGCTCGAAACAGCGGCGTGTATCGT	2155310					
QY	810	GGGTGGACACCTTCGGAAACCCCATGGTGTGCGCGATGCGGAGAAATATGTGGGCGG	869					
DB	2155309	CGGCGGTTCACACTTTTCGTTAAGACCCATGGCG---CGGCGCGGCGCATGTTGGTGGGCC	2155253					
QY	870	AGAGCTGCCGCGCAGGTATTGAAGAAATGAGCCTGGGTGGGAAACACCTACGCGAC	929					
DB	2155252	CGAACCAGAGCTCTCCGCTGGAGAGATGGGGCTGGGTGGAGAGCTCGTATGGGCAC	2155193					
QY	930	CGGACAGGTCGGATACCATCACCACTGAGAGAGCGGCTGGACCAAGACCCCTAC	989					
DB	2155192	CGGAACCGGTAAAGCACCGGATCACCGGATCGAGGTGATGAGACGAACACCCGAC	2155133					
QY	990	TCAATGGAGCAATTAATTTTTTGAACACCTCTTTGTTTACAGTGGGAGCTTACCAAAAG	1049					

QY 2127 AAGACCCCTCGCAGAAAGTGTACCGCTGTGCAGATCTTGAAGAAAAGTTTGTAAAGATT 2186
 Db 2153998 GCGGCGCTTGTGAGGCTATGCGCGCGATGACGCGCAGCGAAGTTCGTGCAGGACTT 2153939
 QY 2167 TGTGAAGCGCTGGCGCAAGTAATGACCTGGACCGGTTTGATCG 2232
 Db 2153938 COTCGCTCGTGGGCAAGGTGATGAACCTCGACAGGTTCGACGTG 2153893

RESULT 8
 ID AAX00818 standard; DNA; 2331 BP.
 AC AAX00818;
 DT 29-MAR-1999 (first entry)
 DE M. tuberculosis catalase peroxide (katG) gene variant.
 KW Catalase-peroxide; katG; mutation; isonicotinic acid hydrazide;
 KW INH; tuberculosis; diagnosis; detection; variant; ss.
 OS Mycobacterium tuberculosis.
 OS Synthetic.

XX Key Location/Qualifiers
 FH CDS 70..2292
 FT /*tag= a
 FT /product= "catalase peroxide variant"
 FT /note= "the start codon is not indicated"
 FT variation 1013
 FT /*tag= b
 FT /note= "G to C variation as compared to the
 wild-type katG gene"

PN W098050585-A1.
 PD 12-NOV-1998.
 XX 06-MAY-1998; 98WO-US09285.
 XX 07-MAY-1997; 97US-0852219.
 XX (MAYO-) MAYO FOUNDATION.
 XX Cockerill FR, Kline BC, Uhl JR;
 PI P-PSDB; AAW95399.
 DR WPI; 1999-070099/06.
 XX
 DR Detection of Mycobacterium tuberculosis - by amplifying katG gene
 and detecting specific fragment, and optionally identifying
 INH-resistant strains by detecting specific mutation
 XX
 PS Claim 3; Page -: 83pp; English.

CC The invention relates to a novel method of detecting Mycobacterium
 tuberculosis. The method comprises amplifying the DNA in the samples to
 generate a detectable amount of amplified DNA comprising a catalase-
 peroxide (katG) DNA fragment with sequence of bases 904-1523 of the
 M. tuberculosis strain H37Rv MC katG gene sequence. The method optionally
 further comprises determining if the katG DNA fragment has a serine to
 threonine mutation in codon 315 (S315T mutation), indicative of an
 isonicotinic acid hydrazide (INH)-resistant M. tuberculosis strain. The
 method can be used to detect M. tuberculosis in biological fluids,
 especially human sputum, useful to diagnose tuberculosis. This disease is
 a major cause of human morbidity and mortality, and conclusive diagnosis
 and subsequent treatment depends on identification of the etiologic agent
 M. tuberculosis. INH has been used in tuberculosis treatment, but
 INH-resistant strains have emerged; the method allows such drug-resistant
 CC strains to be identified. The present sequence represents the M.
 tuberculosis katG gene variant.

CC Note: this sequence is not provided in the specification; it has been
 created by modifying the katG gene sequence given in Fig 7.
 XX
 SQ Sequence 2331 BP; 455 A; 735 C; 762 G; 379 T; 0 other;
 Query Match 37.7%; Score 842.8; DB 20; Length 2331;
 Best Local Similarity 63.0%; Pred. No. 5.8e-253;
 Matches 1389; Conservative 0; Mismatches 802; Indels 15; Gaps 5;

QY 30 TACGTATACACAAACACTGGCGAAATATGCCCTTTTACCGGAGGTTCGCTTAAAGCAAG 89
 Db 93 TACAGAAACCCACCCAGCGCGCTAGCAACGGCTGTCCCGTGGTGGTTCATAGAAATA 152
 QY 90 TGCAGGTGGCGCACCAAAAACACAGGATTTGGTGGCCCAACATCTCAACCTCGGATCTT 149
 Db 153 CCCCGTCGAGGCGCGGAAACACAGGACTGGTGGCCCAACCGGCTCAATCTGAAGTACT 212
 QY 150 AGCCCAACATTCATCGCTATCGGACCCAAACACCGGATTTTGTACTATGCGGAAGATT 209
 Db 213 GCACCAAAACCCGCGCGCTCGCTGACCCGATGGTGGCGTTCGACTATGCGCGGAGGT 272
 QY 210 TAAGAAGCTAGATCTGGCAGCGGTTAAAAAGACCTGGCAGCGCTAATGACAGATTCACA 269
 Db 273 CCGGACCATCGACGTTGAGCGCCCTGACCGGGACATCGAGGAAGTGATGACCACCTCGCA 332
 QY 270 GCACTGGTGGCGCAGCATTTACGGTCATTATATGGCCCTTCTTTATACGCATGGCTGGCA 329
 Db 333 GCGGTGGTGGCGCGCGACTACGGCCACTACGGCGCGCTGTTTATCCGGATGGCTGGCA 392
 QY 330 CAGCCCGCGCCTACCGTATCGGTGATGGCGTGGTGGCGTGGTGGCGTTCACAGCG 389
 Db 393 CCGTCCCGCAGCTACCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 452
 QY 390 CTTCCGCGCTCTCAATAGCTGGCAGACAATGCCAATCTGGATTAAGCAGCGTTCCTTCT 449
 Db 453 GTTCCGCGCGCTTAAACAGCTGGCCCGCAACGCCAGCTTGGACAAGGCGCGCGCGCTGCT 512
 QY 450 TTGCGCCCATCAACAAAATACGGTTCGAAAATCTCCTGGCGGATCTAATGATACTCAC 509
 Db 513 GTGCGCGGTCAAGAAGAGTACGGCAAGAAGCTCTCATGGCGGACCTGNTTGTTCGCG 572
 QY 510 AGGAACCTAGCTCTGGAAACTATGGGCTTTAAACTTTTGGTTTGGAGTGGCGAGCG 569
 Db 573 CGGCAACTGCGCGCTGGAATCGATGGGCTTCAAGACGTTTCGGGTTTCGGCTTCGGCG 632
 QY 570 AGATGATGGGAGCTGAAGAAGATGTATCTGGGGAGCAGAAACCGAATGCTGGGAGA 629
 Db 633 CGACCACTGGGAGCC--CGATGAGTCTATTGGGGCAAGNAGCACCTGGCTCGGCGA 689
 QY 630 CAAGCGCTATGAAGTGACCGAGAGCTCGAAATCCCTGGGAGCGCTACAAATGGGACT 689
 Db 690 TGAGCGTTACAGCGGTAAAGCGGATCTGGAGAACCGCTGGCGCGGTGCAGATGGGCT 749
 QY 690 CATCTATGAACCCCGAAGGACCAACGCGAAGCCACACCTATCGCTGTGCGCGCTGA 749
 Db 750 GATCTACGTGAACCCCGGAGGGGCGCAACCGGACCGCCCATTTGGCGGGCGGTGCGA 809
 QY 750 TATTCGTGAGACTTTTGGCCGAATGGCAATGAATGACGAAGAAACCGTGGCTCTCATAGC 809
 Db 810 CATTCGCGAGAGCTTTCGGCGCATGGCCATGACAGACCTCGAAGACGCGGCGCTGATCGT 869
 QY 810 GGTGGGACACACTTCGGGAAAACCCATGGTGTGCGCATCGGAGAAATATGTGGGCGG 869
 Db 870 CGGCGGTACACTTTCGGTGAAGACCCATGGCGC---CGGCGCGCGGATCTGGTGGGCGC 926
 QY 870 AGAGCTCCCGCGCAGGTATTGAAGAAATGAGCCTGGGTTGGGAAAAACACATCAGGAC 929
 Db 927 CGAAGCCGAGGTGCTCGCTGGAGAGATGGGTTCGGCTTGGAGAGCTCGTATGGGAC 986
 QY 930 CGGACAGCGGTGGGATACCATCACAGTGGACTAGAAGCGCGCTGGACCAAGACCCCTAC 989
 Db 987 CGGAACCGGTAAAGACGGGATCACACCGCGCATCGAGGTCTGTATGGACCAACACCCGAC 1046


```
QY 990 TCAATGGAGCAATAACTTTTGTAAACCTCTTTGGTTACGAGTGGGAGCTTACCAAAAG 1049
Db 1047 GAAATGGGCAACAGTCTTCGAGATCTCTACGGCTACGAGTGGGAGCTGACGAAGAG 1106
QY 1050 TCCAGCTGGAGCTTATCAGTGGAAACAAAGAGAGCGTGGCGGGGTGGCACCATACCGGA 1109
Db 1107 CCGTCTGGCGCTTGGCAATACACCGCAAGAGCGGCGCGGTGCCGCCACCATCCCGGA 1166
QY 1110 TGCACATGATCCCAAGCAAGTGCAGGTCCATTATGCTCACTACGGACCTGGGCGTCG 1169
Db 1167 CCGGTTCG--GCGGGCCAGGCGCTCCCGACGATGCTGCCACTGACCTCTCGTCG 1223
QY 1170 CATGGACCTGATTACGAAATAATTTCTCGAGGTACTATGAAACCTGTAGTGTTC 1229
Db 1224 GGTGGATCCGATATGATGAGCGGATCAGCGCTGCTGGCTGGAACACCCCGAGGATTTGC 1283
QY 1230 AGATGCTTTCCGGAAGCATGTTAACTACACAGAGATATGGACCAAAAGGTGCG 1289
Db 1284 CGACGAGTTCGCCAAGGCTGGTACAGCTGATCACCGACATGGTCCCGTTCCGAG 1343
QY 1290 CTACTGGGACAGAGTGCCTCAGGAAGACCTCATCTGGCAAGACCCCTATACCAAGTGT 1349
Db 1344 ATACTTGGGCGGTGTCCTCAAGCAGACCTGCTGTGGCAGGATCCGCTCCCTGCGGT 1403
QY 1350 AAGCCATCTCTTTAGACGAAACAGATATTTGAAGGCTTAAAGCCAAAATCTCGGAATC 1409
Db 1404 CAGCCACGACCTCTCGGCGAAGCGGAGTTGCCAGCCTTAAGAGCCAGATCCGCGCATC 1463
QY 1410 GGGACTGACGTAAAGCAGGTGGTAAAGCAGGCTATGGCTTCTGCATCTACTTTAGAA 1469
Db 1464 GGGATGACTCTTCACAGCTAGTTTCGACCGCATGGGCGGCGCTGCTGTCCTGG 1523
QY 1470 CTCGACAAGCGCGGTGCCAAGGTGCAGTATACGACTGGCCGCCCAAAAAGACTG 1529
Db 1524 TAGCGACAAGCGCGCGGCCCAAGGTGTCGATCCGCTCGCAGCCACAAGTCGGGTG 1583
QY 1530 GGAAGTAAACACCTCA--GCACTTGCAGGCTACTCAAAACACTAGAGGTATCCA 1586
Db 1584 GGAGGTCAACGACCGGACGCGGATCTCGCAAGGTATTCGACCCCTGGAAGAGATCCA 1643
QY 1587 GGAGGACTTTAACGAGCGGCAATCAGATAACAAAGCAGTATGTTGGCGGACCTGATTGT 1646
Db 1644 GGAGTCATTCACTCCGCGCGCGGGGAACATCAAGTGTCTTCGCGGACCTCGTGT 1703
QY 1647 GCTGGCGGCTGTGGGTGTAGAAAAGCTGCAAAAGATGCTGGCCCATGAGGTGCAGGT 1706
Db 1704 GCTCGGTGCTGTGCGGCATAGAGAAAGCAGCAAGCGCGCTGGCCCAACATCACGGT 1763
QY 1707 GCCTTTCAACCGGACGAGCGGATGCCACGCTGAGCAAAACCGATGTGGAAGCTTCGA 1766
Db 1764 GCCCTTCAACCGGCGCGCATGCGTGCAGGAACAAACCGGACGTTGAATCCTTTGC 1823
QY 1767 AGCACTAGAGCCAGCGGCTGACGGCTTTAGAAACTACATTAACCGGAGCAATAAGTATC 1826
Db 1824 CGTGTGAGGCCAAGGAGATGCTTCCGAACTACTCCGAAGGCAACCCGTTGCC 1883
QY 1827 CGGTGAGGAATGTCTGTAGACCGGCGGCGAGCTTCTGTGCTGGCCACCAAGAAATGAC 1886
Db 1884 GGCGGAGTACATGTGCTGCGAAGCGCAAGCGCTGCTTACGCTAGTGCGCCCTGAGATGAC 1943
QY 1887 TGCCTTTGTAGCGGTATGCTGTACTGGCGACCAACTACGACGTTTCGACATGGAGT 1946
Db 1944 GGTGCTGTAGGTGGCTTGGCGGCTCTCGGGCAAACTACAGCGCTTACCGCTGGGCGT 2003
QY 1947 GTTTACAAATAAGCGGCTCAGCTATCCAATGACTCTTTGTAAACCTGCTAGACCTCAA 2006
Db 2004 GTTACCGGCGCTCCGAGTCACTGACCAACGACTTCTGCTGAACCTGCTCGACATGGG 2063
QY 2007 CACTAAATGGGAGCCAGCGATGAATCAGCAAAAGTTTTTGAAGCGAGAGCTTCAAAAC 2066
Db 2064 TATCACCTGGGAGCCCTCGCCACGAGATGACGGGACCTACCAGGGCAAGGATGGC--AG 2120
QY 2067 TGGCGAAGTAAGTGGAGTGGCACCGCGGTAGACCTGTATCTTCGATCCAAATTCGAGCT 2126
```

RESULT 9

AAT89967

ID AAT89967 standard; DNA; 2235 BP.

XX AAT89967;

XX 20-MAR-1998 (first entry)

XX Mycobacterium tuberculosis partial katG gene.

XX Tuberculosis: katG gene; isoniazid resistance; INH;

XX isonicotinic acid hydrazide; ss.

XX Mycobacterium tuberculosis.

FH Key Location/Qualifiers

CDS 85..2235

FT /*tag= a

FT /product= "katG gene"

FT /note= "partial cds"

XX US5688639-A.

XX 18-NOV-1997.

XX 18-APR-1994; 94US-0228662.

XX 18-APR-1994; 94US-0228662.

XX (MAYO-) MAYO FOUNDATION.

XX Cockerill FR, Kline BC, Uhl JR;

XX WPI; 1998-007975/01.

XX Determination of isoniazid sensitivity of Mycobacterium tuberculosis strains - by restriction length polymorphism analysis of katG gene

XX Claim 1; Column 9-12; 18pp; English.

XX This partial DNA sequence encodes the katG gene of Mycobacterium tuberculosis strain H37Rv which is used in a novel method to rapidly identify strains of M. tuberculosis which are resistant to isoniazid (INH, isonicotinic acid hydrazide). The method involves the use of restriction fragment length polymorphism (RFLP) analysis to determine if a NciI-MspI restriction site is absent in the DNA of the strain at the codon corresponding to codon 463 of a M. tuberculosis katG gene consensus sequence (see AAW31343). The absence of the site indicates an INH-resistant strain.

XX Query Match 37.2%; Score 832.4; DB 19; Length 2235;

XX Best Local Similarity 63.0%; Pred. No. 1e-249;

XX Matches 1389; Conservative 0; Mismatches 801; Indels 16; Gaps 6;

QY 30 TACGTATACACAACTGCGGAAATGCCCTTTTACCGGAGGTTCGCTTAAGCAAAG 89

Db 33 TACAGAAACCAACCCGCGGCTAGCAACGGCTGTCCCGTCTGGGTTCATATGAATA 92

QY 90 TGCAGGTGGCGGCAACCAAAACAGGATTTGGTGGCCCAACATGCTCAACCTCGGCATCTT 149

[illegible]

QY	330	CAGCGCCGGCACCTTACCGTATCCGGTGATGGCCGCTGGTGGCGGTGGCTCCGGCTCACAGCG	389
Db	333	CGTTCGCGGGACCTTACCGCATCCACGACGCGCGCGGGCGCGCATGCAGCG	392
QY	390	CTTCGCGCCTCTCAATAGCTGGCCAGACAAATGCCAATCTGGATAAAGCACGCTTGCTTCT	449
Db	393	GTTTCGCGCGGCTTAACAGCTTGGCCGCAACGCGCAGCTTGGCAAGCGCGCGCGGTCT	452
QY	450	TTGCGCCATCAACAAAAATACGGTCGAAAAATCTCTCTGGCGGATCTAATGATACTCAC	509
Db	453	GTGCGCGGTCAAGAGAAGTACGCGAAGAGCTCTATGGCGGACGCTGATGTTTTCG	512
QY	510	AGGAAC--GTAGCTCTGGAACATATGGGCTTTAAAACTTTGGTTTTCAGGTGGCAGA	567
Db	513	CGGCAACGGCTCGCCTCGGAATCGATGGGCTTCAAGACGTTTCGGGTTTCGGCTTCGGCGT	572
QY	568	GCAGATGTATGGGAGCTGAAGAAGATGTATCTGGGGACAGAAAACCGAATGGCTGGGA	627
Db	573	CGACGAGTGGGAGACC-----GATGAGGTCTATTGGGCAAGGAAGCAACCTGGCTGGC	627
QY	628	GACAAGCGCTATGAAGGTGACCGAGAGCTCGAAAAATCCCTCTGGGAGCGTACAAATGGGA	687
Db	628	GATGA---CGGTTACAGCGTTAAGCGATCTGGAGAACCCGTGGCCCGGTGTCAGATGGGG	684
QY	688	CTCATCTATGTAAACCCCGGAGGACCAACGGAAGCAGACACCTATCGCTGTGTGGCGT	747
Db	685	CTGATCTACGTGAACCCGGAGGCGCCAGCGCAACCCGACCCCATGGCCGCGCGCGGTC	744
QY	748	GATATTCTGTGAGACTTTTGGCCGNAATGGCAATGNAATGACGAAGAACCGTGGCTCTCATYA	807
Db	745	GACATTTCGGCAGAGCTTTTCGGCGCATGGCCATGAACGACGTCGAAACAGACGGCGCTGATC	804
QY	808	CGCGGTGGACACACTTTCGGAATAACCCATGTGTCTGCCGATGCGGAGAAATATGTGGGC	867
Db	805	GTCCGCGGTACACTTTCGGTTAGACCCATGG---CGCCGCGCCGCGCGATCTGGTCGGC	861
QY	868	CGAGAGCCTGCCGCCGAGGTATTGGAAGAATGAGCCTGGGGTGGAAAAACACTACGCG	927
Db	862	CCCGAACCCGAGGTGCTCCGCTGGAGCAGATGGGCTTGGGCTTGAAGAGCTCTGATGCG	921
QY	928	ACCGGACACGTTGGGATACCATCACAGTGGACTAGAAAGCGCGCTGGACCAAGACCCCT	987
Db	922	ACCGGAACCGGTAAAGACGGGATCACACGGGATCGAGTCTGTATGGACGAACACCCCG	981
QY	988	ACTCAATGGAGCAATAACTTTTGAANAACCTCTTTGGTTACGAGTGGGAGCTTACCAA	1047
Db	982	ACGAAATGGGCAACACATTTCTCTCGAGATCTCTGTACGGCTACGAGTGGGAGCTGACGAAG	1041
QY	1048	AGTCCAGCTGGAGCTTATCAGTGGAAACCAAAAGACGGTCCGGGGCTGGACCATACCG	1107
Db	1042	AGCCCTGCTGGCGCTTGGCAATACCCGCGCAAGAGCGCGCGGTGCCGSCACCATCCCG	1101
QY	1108	GATCCACATGATCCACACAAGTCGCAGCTTCCATTTATGCTCACTACGGACACTTGGCGCTG	1167
Db	1102	GACCCGTTCG---GCGGGCGACGGCGCTCCCGACGATGCTGGCCACTGACCTCTCGCTG	1158
QY	1168	CGCATGACCCTGATTACGAAAAAATTTCTCGAGGTTACTATGNAACCCCTGATGAGTTT	1227
Db	1159	CGGTTGATCCGATCTATGAGCGGATCACCGCTCGCTGGCTGGAAACCCCGAGGAATTG	1218
QY	1228	GCAGATGCTTTCCGGAAGCATGGTCAAAACTGACACACAGATATGGGACCAAAAGTG	1287
Db	1219	GCCGACGATTCGCAAGGCTTGGTACAAGCTGATCCACCGACATGGGTCCCGTTGCG	1278
QY	1288	CGTACCTGGGACGAAGTGGCTTCAGGAAGACCTCATCTTGGCAAGACCCATATACCAT	1347
Db	1279	AGATACCTTTGGCGGCTTGGTCCCAAGCACACCTGCTGTGGCAGGATCCGGTCCCTGCG	1338
QY	1348	GTAAGCCATCTCTTTAGAGAAACGATATTGAAGGCTTAAAGCCAAATTCCTGGAA	1407
Db	1339	GTACG---ACGACCTCGTCCGGGAAGCAGATTGCCAGCTTAAAGACCGAGATCCGGGA	1395

1408	TCGGGACTGACGTTAGCGAGCTGGTAAGCACCGCATGGGCTTCTGCATCTACTTTTAGA	1467
Qy		
1396	TCGGGATTGACTTCTCACAGCTAGTTTTCACACCGCATGGCGCGCGCTGCTGTTCCGT	1455
Db		
1468	AACCTCTGACAAGCGCGCGGTGCAACGGGTGACGTTATACGACTGGCCCCACAAAAGAC	1527
Qy		
1456	GGTAGGACANGCGCGCGCGCAACGGGTGGTCATCGCCTGCAGGCCACAAGTCGGG	1515
Db		
1528	TGGGAAGTAACAACCCCTCAGCAACTTGGCCAGGGTACTCAAAACACTAGAAAGTATCCAG	1587
Qy		
1516	TGGGAGTCAACGACCCGACGGGATCTGCGCAAGGTCATTTGCGCACCTCTGAAGAGATCCAG	1575
Db		
1588	GAGGACTTTAAACAGGGGCAATCAGATAACAAGAAGTAGTATCGTTGGCCGACCTGATTGTG	1647
Qy		
1576	GAGTCATTTCA-----CTCGCGCGGGAACATAAAGTGTCTTTCGCCGACCTCGTCGTG	1629
Db		
1648	CTGGCGGCTGTGCGGTGTAGAAAAAGCTGCAAAAAGATGCTGGCCATGAGGTGCAAGTG	1707
Qy		
1630	CTCGGTGGCTGTGGCCACTAGAGAAAGCAGCAAAAGCGGCTGGCCACAACATCAGCGTG	1689
Db		
1708	CTTTTCAACCCCGGAGAGCGGATGCCCGCTGAGCAAAACCGATGTGGAAAGTTCGAA	1767
Qy		
1690	CCCTTCAACCCCGGCGCACGATGGCTCGCAGGAACAACCGACGCTGGAATCCTTTGCC	1749
Db		
1768	GCCTAGAGCCAGCGCTGACGGCTTTAGAACTACATTAACCCGGAGCATAAAGTATCC	1827
Qy		
1750	GTGCTGGAGCCCAAGCGAGATGGCTTCCGAAACTACTCGGAAAGGCAACCGTTGCCCG	1809
Db		
1828	GCTGAGGAATGCTCGTAGACGGGCGCAGCTTCTGTGCTTTTCGCCACACAGAAATGACT	1887
Qy		
1810	CCGAGTACATCGTCTCGACAAGGCGNACTGCTTACGCTCAGTGCCTTGAGATGACG	1869
Db		
1888	GCCTTGGTAGGGCGGTATGCGTGTACTGGGCACCAACTCAGACGGTTTCGCAGCATGAGTG	1947
Qy		
1870	GTGCTGTTAGTGGCTTCGCGCTCTCGGCGCAAACTACAAGCGCTTACCGCTGGCGGTG	1929
Db		
1948	TTTACAAATAAGCCGGGTGAGCTATCCAATGACTTCTTGTAACTGCTAGACCTCAAC	2007
Qy		
1930	TTCAACGAGCCCTCCGAGTCACTGACCAAGCACTTCTTGTGAACCTGCTCGACATGGGT	1989
Db		
2008	ACTAAATGGCGACCCAGCGATGAATCAGACAAAGATTTTGAAGCGACAGACTTCAAAACT	2067
Qy		
1990	ATCACCTGGGAGCCCTCGCCAGCAGATGACGGGACCTACCAGGGCAAGGATGSC---AGT	2046
Db		
2068	GGCGAAGTAAAGTGGAGTGGCACCCGGGTAGACCTGATCTTCGGATCCAAATCCCGAGCTA	2127
Qy		
2047	GGCAAGGTGAAGTGGACCGGCGACCGCGTGGACCTGTGCTTCGGGTCCAACCTCGGAGTTG	2106
Db		
2128	AGAGCCCTCCGCAAGGTGTACGGCTGTGCAGATTCTGAAGAAAAGTTTGTAAAGATTTT	2187
Qy		
2107	CGGGCGCTTGTCCAGGTCTATGCGCCGATGACCGGCGAGGTTCTGTGACAGGATTC	2166
Db		
2188	GTGAAGGCTTGGGCCAAAAGTAATGGACCTGGACCGGTTTGATCTG	2232
Qy		
2167	GTCCGCTCGCTGGACAAAGGTGATGAACCTTCGACAGGTTGCACGTG	2211
Db		

RESULT 11

Accession ID	Standard	DNA	BP
AAQ51531	standard	DNA	4795
XX	AAQ51531		
XX	AC		
XX	AC		
XX	AC		
DT	25-MAR-2003	(updated)	
DT	17-MAY-1994	(first entry)	
XX			
XX			
DE	M. tuberculosis H37Rv	katG gene.	
XX			
XX			
KW	catalase-peroxidase; isonicotinic acid hydrazide;	isoniazid;	INH;
KW	antibiotic; susceptibility; sensitive;	katG;	ss.
XX			
OS	Mycobacterium tuberculosis	(strain H37Rv).	
XX			

```
Db 2302 CGTGGCGGACCTACCGCATCCACGAGCGCCGGCGGCGCCGGGGCGGCATGCAGCG 2361
QY 390 CTTGCGGCCCTCAATAGCTGGCCAGACAAATGCCAATCTGGATAAAGACAGCTTGCTTCT 449
Db 2362 GTTCGGCGCCGTTAAACAGCTGGCCGACAAACGCGAGCTTGGACAAGCGCGCGCTGCT 2421
QY 450 TTGGCCCATCAACAAATAACGGTGCAGAAATCTCTGGCGGGATCTAATGATPACTCAC 509
Db 2422 GTGGCCGGTCAAGAAGAGTAGGCAAGAGCTCTCATGCGCGGACCTGATTTTTCGCG 2481
QY 510 AGGAAAC--CTAGCTCTGGAACATATGGGCTTTAAACATTTTGTGTTTTCAGGTGGCAGA 567
Db 2482 CGCAACCGCTGCGCTCGGAATCGATGGGCTTCAAGACGTTTCGGGTTTCGGGCGT 2541
QY 568 GCAGATGATGAGAGCTGAAGAAGATGTATATCTGGGAGCAGAAACCCGAATGGCTGGGA 627
Db 2542 CGACCAGTGGAGACC-----GATGAGTCTATTGGGGCAAGGAAGCCACCTGCTCGCG 2596
QY 628 GACAAGGCTATGAAGGTGACCAGAGCTCGAAATCCCTGGAGCCGTACAAATGGGA 687
Db 2597 GATGA---CGGTTACAGCTGAAGCATCTGGAAACCCGCTGGCCGGGTGCAAGATGGG 2653
QY 688 CTCATCTATGAAACCCGCAAGCAACGCAAGCCAGACACCTATCGCTGCTGCGCGT 747
Db 2654 CTGATCTAGTGAACCGGAGGCGCGCAACGCGCAACCCGACCCCATGGCCGCGCGGTC 2713
QY 748 GATATTCGTGAGACTTTTGGCCGAATGGCAATGAATGACGAAGAAACCCGTGGCTCTCATYA 807
Db 2714 GACATTCGCGAGAGCTTTTCGGCGCATGGCCATGAACGACGTCGAAACAGCGCGCTGATC 2773
QY 808 GCGGGTGGACACACTTCGAAACCCCATGGTGTGCCATCGCGGAGAAATATATGGGC 867
Db 2774 GTCGGGGTCACACTTTTCGTTAGACCCCATGG---CGCCGGCCGCGCGCATCTGGTGGC 2830
QY 868 CGAGAGCTCCCGCGCAGGTATTGAAGAAATGAGCTGGGTGGAAACACACTACGCG 927
Db 2831 CCGGAACCCGAGGCTGCTCCGCTGGAGCAGATGGCTTGGGCTTGGAGAGCTCGTATGGC 2890
QY 928 ACCGGACACGGTGGGATACCATCACAGTGGACTAGAAGCGCGCTGGACCAAGACCCCT 987
Db 2891 ACGGGAACCGGTGAAGCAGCGCATCACGAGCGCATCGAGGTCGTATGGACGAACACCCCG 2950
QY 988 ACTCAATGGAGCAATACTTTTGAACCTCTTTGTTTACAGTGGGAGCTTACCAAA 1047
Db 2951 ACGAATGGGACACAGTTTCTCTCGAGATCTGTACGGCTACGAGTGGGAGCTGACGAAG 3010
QY 1048 AGTCCAGCTGGAGCTTATCAGTGGAAACCAAGACGGTCCCGGGGTGGCAACATACCG 1107
Db 3011 AGCCCTGCTGGCGTTGGCAATACACGGCCAAGGACGGCGCGCTGCCGACCATCCG 3070
QY 1108 GATGCACATGATCCACGAGTGCAGGCTCCATTTATGCTACTAGGAGCTGGCGGTG 1167
Db 3071 GACCCGTTGC---GCGGGCCAGGCGGCTCCCGACGATGCTGGCCACTGACCTCTCGCTG 3127
QY 1168 CGCATGACCCCTGATTAACGAAAAATTTCTCGAGGTAATGAAACCCCTGATGAGTTT 1227
Db 3128 CGGGTGGATCCGATCTATAGCGGATCACCGCTGCTGGCTGGAACACCCCGAGGAATTG 3187
QY 1228 GCAGATGCTTTTCGGAAGCATGGTACAAACTGACACACAGAGATATGGACCAAAAGGTG 1287
Db 3188 GCGGACGAGTTCCGCAAGGCTGTGTACAAGCTGATCCACCGAGACATGGGTCCGTTGCG 3247
QY 1288 CGGTACTCTGGGACAGAGTGGCTTCAGGAAGAGCTCATCTGGAAGACCCCTATACACAGAT 1347
Db 3248 AGATACCTTGGCGGCTGGTCCCAAGCAGACCTGCTGTGGCAGGATCCGGTCCCTGCG 3307
QY 1348 GTAAGCATCTCTTGTAGACGAAGACGATTTGAAGGCCCTAAAGCCAAATCTCTGGAA 1407
Db 3308 GTCAGC---ACGACCTGCTGGCGGAAGCAGATTGCCAGCCTTAAGAGCCAGATCCCGGCA 3364
QY 1408 TCGGAGCTACGGTGAAGCGAGCTGGTGAAGCAGCGCATGGCTCTGCATCTACTTTTGA 1467
```

```
Db 3365 TCGGATTGACTGTCTCACAGCTAGTTTCGACCCGATGGGGGGGGGCGTCTGTCGTTCCGT 3424
QY 1468 AACTCTGACAAAGCGCGCGTGCACACGTCACAGTATAGACTGGCCCCCACAACAAAGAC 1527
Db 3425 GGTAGCGAAGCGCGCGCGCAACGGTGGTGCATCCGCTGCAGCCACACAGTGGG 3484
QY 1528 TGGGAAGTAAACAACCTCAGCAACTTGCAGAGGTACTCAAAACACTAGAAAGTATCCAG 1587
Db 3485 TGGGAGGTCAACGACCCCGCGATCTGCGCAAGGTCAATTCGCAACCTGAAGAGATCCAG 3544
QY 1588 GAGGACTTTTAACGAGCGCAATCAGATAACAAGACGATATCGTTGGCGCGACCTGATTGTG 1647
Db 3545 GAGTCATTCA-----CTCGCGCGGGGAACATCAAAAGTGCTTCGCGGACCTCGTGTG 3598
QY 1648 CTGCGCGGCTGTGCGGCTGTAGAAAAAGCTGCAAAAGATGCTGCCCATGAGGTGAGGTG 1707
Db 3599 CTGCGTGGCTGTGCGCACTTAGAGAAAGCAGAAAGCGGTGCGCCACAACATCACGGT 3658
QY 1708 CCTTCAACCGCGGACGCGGATGCCACCGCTGAGCAAAACCGGATGTGGAGCTTTTCGAA 1767
Db 3659 CCCTTCAACCGCGCGCGGCGGATGCGTGCAGGAACAAACCGACGTGGAATCCTTTGCC 3718
QY 1768 GCATGAGCAGCGCGCTGACGGCTTTAGAAACTACATTAACCGGAGCATAAAGTATCC 1827
Db 3719 GTGCTGAGCGCCAAAGCAGATGGCTTCGGAACACTACCTCGAAAGGCAACCGTTGCCGG 3778
QY 1828 GCTGAGGAATGCTCTGAGACCGCGCAGCTCTGTCGCTTTTCGGCACCAGAAATGACT 1887
Db 3779 CCGAGTACATCGCTGCTCGACAAAGCGGACCTGCTTACGCTCAGTGCCTGAGATGACG 3838
QY 1888 GCCTTGTAGCGGATGCTGTACTGGGCACCAACTACGACGGTTTCGAGCATGGAGTG 1947
Db 3839 GTGCTGTGATGGCTGCGCGCTCTCGCGCAAACTACAAGCGCTTACCGCTGGGCGTG 3898
QY 1948 TTTACAATAAGCGCGGTCAGCTATCAATGACTTCTTTGTAAACCTGCTAGACCTCAAC 2007
Db 3899 TTCAACGAGGCTCCGAGTCACTGACCAACGACTTCTTCGTGAACCTGCTCGACATGGT 3958
QY 2008 ACTAAATGGGAGCCAGCGATGAATCAGACAAAGTTTTTGAAGCAGAGACTTCAAAACT 2067
Db 3959 ATCACTTGGGAGCCCTCGCCAGCAGATGACGGGACCTACCAAGGCAAGGATGGC---AGT 4015
QY 2068 GCGCAAGTAAAGTGGAGTGGCACCGCGGTAGACCTGATCTTCGATCCAATTCGAGCTA 2127
Db 4016 GGCAGGTGAAGTGGAGCGCGCGCTGAGCTGTCTTCGGGTCCAACTCGGAGTTG 4075
QY 2128 AGAGCCCTCCAGAAAGTGTACGGCTGTGAGATTCTGAAGAAAAATTTGTTAAAGATTTT 2187
Db 4076 CCGCGCTTGTGAGGTTCTATGCCCGGATGACGCGGCGAGGCAAGTTCGTGACAGGATTC 4135
QY 2188 GTGAGGCCCTGGGCAAAAGTAATGGACCTGGACCGGTTTGAATCTG 2232
Db 4136 GTCGCTGCTGGGCAAGGATGAACCTCGACAGGTTTCGACGTG 4180
```

RESULT 12

AA16969 standard; DNA; 4795 BP.

AA16969;

20-MAR-2003 (updated)

11-MAY-1999 (first entry)

M.tuberculosis katG gene.

katG; catalase; peroxidase; enzyme; isonicotinic acid hydrazide;
isoniazid; INH; diagnosis; resistance; ss.

Mycobacterium tuberculosis.

Key Location/Qualifiers
CDS 1979..4186

QY 1588 GAGGACTTTAACCGCGCGCAATCAGATACAAAGCAGTATCTGTGGCGGACCTGATGTG 1647
 DB III III III III III III III III III III III III III III III III
 3545 GAGTCATTCA-----CTCGGCGCGGGAACATCAAAGTGCTCTCGCCGACCTCGTCGTG 3598
 QY 1648 CTGGCGGCTGTGGGGGTAGAAAAGCTGCAAAAGATGCTGGCCATGAGCTGCGAGTG 1707
 DB III III III III III III III III III III III III III III III III
 3599 CTCGGTGCTGTGCGGCCTAGGAAGAGCAGAAAGCGGCTGGCCACAAATCATCAGGTG 3658
 QY 1708 CTTTCAACCCCGGAGCGGATGCCACCGCTGAGCAACCGATGTGGAACTTTTCGAA 1767
 DB III III III III III III III III III III III III III III III III
 3659 CCTTCAACCCCGGCGCGCAGCATGCTGCGAGGAACAAACGAGTGGAACTTTCGTC 3718
 QY 1768 GCACATAGACCGAGCGCTGACCGCTTTAGAAATCATTAACCGGAGCATTAAGATGCC 1827
 DB III III III III III III III III III III III III III III III III
 3719 GTGCTGGAGCCCAAGCGAGATGCTTCGGAATACCTCGGAAGGCGCAACCGTTGCCGG 3778
 QY 1828 GCTGAGGAAATGCTGCTAGACCGGCGCAGCTTCTGTGCTTTTCGCGACCAAGAAATGACT 1887
 DB III III III III III III III III III III III III III III III III
 3779 CCGAGTACATGCTGCTGCACAAGGGGAACCTGTACGCTCAGTGCCCTGAGATGAGC 3838
 QY 1888 GCTTTGGTAGCGGTATCGGTACTGTGGCAACCACTAGACGGTTTCGCGAGATGAGTG 1947
 DB III III III III III III III III III III III III III III III III
 3839 GTGCTGGTAGTGGCTGCGCTCTCGCGCAAACTACAAGCGCTTACCGCTGGGCGTG 3898
 QY 1948 TTTACAATAACCGGCTGAGTATCCATGACTTCTTTGTAAACCTGTAGACCTCAAC 2007
 DB III III III III III III III III III III III III III III III III
 3899 TTCACGAGGCTCGCGAGTCACTGACCAACGACTTCTGCTGAACTGCTGACATGGGT 3958
 QY 2008 ACTAATAGCGGAGCGAGCATGAATCAGACAAAGTTTTTTGAAGGCGAGACTTCAAACT 2067
 DB III III III III III III III III III III III III III III III III
 3959 ATCACTGGAGCCCTCGCCACGATGACGGGACCTACCGAGGCAAGGATGGC---AGT 4015
 QY 2068 GCGAAGTAAGTGGAGTGGCGACCGCGGTAGACCTGATCTTCGGATCCAAATTCGAGCTA 2127
 DB III III III III III III III III III III III III III III III III
 4016 GCGAAGTGAAGTGGACCGGCGCGGTGGACCTTCTCGGGTCCAACTCGGAGTTG 4075
 QY 2128 AGAGCCCTCGAGAGTGTACGGCTGTGCGAGATCTGGAAGAAAGTTTGTAAAGATTTT 2187
 DB III III III III III III III III III III III III III III III III
 4076 CGGGCGCTTGTGAGGTCTATCGCGGATGACGCGCGAGCGGCAAGTTCTGACAGATTC 4135
 QY 2188 GTGAAGGCTGGCGCAAGTAATGACCTGGACCGGTTTGTGATCTG 2232
 DB III III III III III III III III III III III III III III III III
 4136 GTCGCTGCTGGGACAAAGTGATGAACCTCGACAGGTTTCGACGTG 4180

RESULT 13
 AAQ06815
 ID AAQ06815 standard; DNA; 2196 BP.
 XX AC AAQ06815;
 XX

DT 06-MAR-1991 (first entry)
 XX DE Sequence encoding heat resistant peroxidase.
 XX

KW pod10; E.coli UM228; ds.
 XX OS Bacillus stearothermophilus.

PH Key Location/Qualifiers
 FT CDS 1..2193
 FT /*tag= a

XX JP02268684-A.

XX PD 02-NOV-1990.

XX PF 07-APR-1989; 89JP-0089469.

XX PR 07-APR-1989; 89JP-0089469.

XX PA (TOYM) TOYOB KK.

XX WPI; 1990-372008/50.
 DR P-PSDB; AAR08205.
 XX DNA contg. genetic information of heat resistant peroxidase -
 PT prepd. by culturing transformant and collecting heat resistant
 PT peroxidase
 XX Disclosure; Fig 2; 8pp; Japanese.
 XX Sequence may be used to construct plasmid pod10 used to transform
 CC E.coli UM228. The transformed expression system may be used to
 CC produce heat resistant peroxidase in large quantities.
 XX Sequence 2196 BP; 576 A; 564 C; 633 G; 423 T; 0 other;
 Query Match 32.2%; Score 721.6; DB 11; Length 2196;
 Best Local Similarity 62.7%; Pred. No. 5.7e-215;
 Matches 1336; Conservative 0; Mismatches 754; Indels 42; Gaps 12;
 QY 101 GCACCAAAACAGGGATGGTGGCCCAACATGCTCAACCTCGGCATCTTACGCCAACATT 160
 DB III III III III III III III III III III III III III III III III
 74 GAACGACGAACAAAGACTGTGGCCGAACAGCTGAACCTTAAGCATTTCCATCAACATG 133
 QY 161 CATCGCTATCGGACCCAAAGACCCGGATTTTGTACTATGCGGAAAGAGTTTAAGAAGCTAG 220
 DB III III III III III III III III III III III III III III III III
 134 ACCGAAAAACGAATCCTCATGATGAAGAGTTCAACTATGCTGAGGAGTTTCAAAAACTAG 193
 QY 221 ATCTGCGACGGTTAAAAAGGACCTGGCAGCGCTAATGACAGATTACAGAGCTGCTGGC 280
 DB III III III III III III III III III III III III III III III III
 194 ACTATTGGCGCTCAAGAAGATTTGCGCAACTGATGACGGAACCAAGACTGCTGGC 253
 QY 281 CAGCAGATTACGCTGCTATTATGCCCCCTTCTTTATACGCATGGCTGCACAGCGCCGCA 340
 DB III III III III III III III III III III III III III III III III
 254 CGGCCGATTATGCCATTACGGCCGCTGTGTTATCCGATGGCTGGCATTCAGCTGGCA 313
 QY 341 CTTACCGTATCGGTGATGCGCGTGGTGGCGGTGCTCGGCTCAGACGGCTTCGCGCTC 400
 DB III III III III III III III III III III III III III III III III
 314 CGTACCGCATCGCGCGCGCGCGCGCTTCGACCGCGCAGCGCGCTTTGCGCGCT 373
 QY 401 TCAATAGCTGGCGAGACAATGCCAATCTGATATAAGCAGCTTGTCTTTGGCCCATCA 460
 DB III III III III III III III III III III III III III III III III
 374 TAAACAGCTCGCGGACACGCGCACTTGGATAAAGCGCG- GCGTTGTTATGCGCCATCA 432
 QY 461 AAAAAAATACGCTCGAAAAATCTCTCGGCGGATCTAATGACTACTCAGAGAAACCTAG 520
 DB III III III III III III III III III III III III III III III III
 433 AAAAGAAATACGGGAACAAAATCTCTTGGCGCGATTTGTTCAATTTTGGCGCAATGTCG 492
 QY 521 CTCTGAAAATATGCGCTTTAAACCTTTTGGTTTTCGAGTGGCAGAGCAGATGTATGG 580
 DB III III III III III III III III III III III III III III III III
 493 CTATTGAATCGATGGGTGGAAAAACGATTTGGATTCGCGCGCGCGCGCTTTGAGCTGCG 552
 QY 581 AGCCTGAAGAAGATGTATCTGGGAGCAGAAACCGAATGGCTGGGAGACAACGCGTATG 640
 DB III III III III III III III III III III III III III III III III
 553 ATCCGGAAGAAGACGTTTATTTGGGATCGGAAAAAGAGTGCTGCGCTCTGAACGCTATT 612
 QY 641 AAGGTGACCGAGAGCTCGAAAAATCCCTCGGAGCCCTTACAAATGGGACTCATCTATGTA 700
 DB III III III III III III III III III III III III III III III III
 613 CCGGTGATCGGACCTCGAAAAACCCGCTCGCCG- CGTGCAAAATGGGTTAATCTAGTCA 671
 QY 701 ACCCGAAGACCCCAACGGCAGCCAGACCTTATCTGCTGCTGCGCGTGTATTCGTGAGA 760
 DB III III III III III III III III III III III III III III III III
 672 ACCCAGAAGGGCGGACGCGGAGATCCAAAAGCAGGGGATATCCGCA----GAGA 727
 QY 761 CTTTTCGCGGAATGCAATGAATGACGAAGAACCGTGGCTCTCATAGCGGTGGACACA 820
 DB III III III III III III III III III III III III III III III III
 728 CGTTCGCGCATGGGGATGAAGATGAAGAAACGGTTCGCTTGTATCGCGCGGCTCAT 787
 QY 821 CTTTCGGAATAACCCATGCTGCTGCGGATGCGGAGAAATATGTGGCGGAGAGAGCTGCC 880
 DB III III III III III III III III III III III III III III III III
 788 CGTTCGGAATAACCGCACCGCGCGCGCTCC----CAGGCACGTCGTCGCCGCGGAG 844
 QY 881 CCGCAGGTATTGAAGAAATGAGCCTGGGGTGGAAAAACACCTACGGAACCGGACGCGTG 940

	99	150	159	210	219	270	279	330	339	390	399	450	459	510	519	570	579	630	633	690	684	750	744	810	804	870	864	927	924	987	984	1047	1044	1107	1104																																			
QY	CGCACCAAAACAGGAGATTGGTGGCCCAACATGCTCAACCTCGGATCTTATAGCCAAACA	158	CGGTACTTCCACAAAGATTGGTGGCCGAAGGTTGCAACCTGGATTTTGGATCAGCA	209	TTTCATCGCTATCGGACCCAAACAGCACCGGATTTTGTAGCTATGCGCAAGAGTTTAAAGA	218	AGATCGCAATCAGACCCGATGGATCCGGATTTCACTACCGTGAAGAAGTACGCAAGCT	269	AGATCTGGCAGCGCTTAAAAAGACACCTGGCAGCGCTAAATGACAGATTTCACAGGACTGGT	278	CGATTTTCGACGCGCTGAAGAAGATGTCACGCGTTGATGACCGCATACGCAAGAGTGGT	329	GCACGAGATTACGGTCATTTATGGCCCTTCTTTATACGATGCGCTGGCAGCAGCGCG	338	GCCCGCTGACTGGGGGACATACGCGGTTTGTATCCGTATGGCTTGGCATCCGCTGG	389	CACCTACCGTATCGGTGATCGCGCTGGTGGCGTGGCTCCGGCTCAGCGCTTCGCGCC	398	CACCTACCGTATTCGCTGATGGCCGTTGGGGCGGTGGTACCGGAAGCAGCGCTTTGCACC	449	TCTCAATAGCTGGCCAGACAATGCCAATCTGGATAAAGCACGCTTGCTTCTTTGGCCCAT	458	GCTCAACTCTGGCCGACAAAGTCAAGCTGGATAAAGCGCGCGCTGCTGTGGCCGAT	509	CAAAACAAAATACGGTCGAAAAATCTCTTGGCGGATCTTAATGATATCTACAGGAACGT	518	CAAGAAGATACGGCAACAAAATCAGCTGGCAGACCTGATGATTTCTGGCTGGCACCGT	569	AGCTCTGGAACATGGGCTTTAAACCTTTTGGTTTTTGCAGGTGGCAGACGATGTATG	578	GGCTTATGATCCATGGGCTTACTGCTTACGGCTTCTCTTTCGGCCGCGTCGATATTG	629	GGAGCCTGGAAGAAGATGATACTGGGAGCAGAAACCGAATGGCTGGGA-----GACAA	632	GAAACCGAAAAAGATATCTACTGGGCTGACGAAAAAGAGTGGCTGGCACCTTCTGACGA	689	CGGCTATGAGGTGACCGGAGC-----TGGAATATCCCTGGGAGCCGTACAAAT	683	ACGCTACGGCGACGTGAACAAACGACAGACCATGGAAGACCGCTGGCGCTGTCCAAAT	749	GGGACTCATCTATCTAAACCCGAAGACACCAACGCAAGCAGACCTATCGCTGTGTC	743	GGGTCTGATCTATGTGAACCCGAGAGTGTTAAGGGCCACCGCTGATCGCTGGAACCGC	809	CGGTGATATTGCTGAGACTTTTGGCCGAATGGCAATGAATGACGAAGAACCGTGGCTCT	803	ACAGCAGGTACTTTGAAACCTTCGCCCGTATGGCGATGAAGCAGCAAAAAACCGCAGCCT	869	CATAGCGGTGGACACACCTTCGGAAAAACCCATGGTGTGCTGCCGATCGGGAAGATATGT	863	CACAGCTGGCGGCGACACCGCTCGGTAAATGTACGGTAAATGGCAATGCCCTGTGCGT---	926	GGGCCGAGACGCTGCCCGCAGGTATTGAAGAAATGAGCTGGGGTGGAAAAACACTA	923	AGCCCTTGACCCAAAAGCCCTGACGTTGAAAACCCAGGGCTTAGTTGGGGCAACCCCAA	986	CGGACCGGACACGGTGGGATACCATCACAGTGGACTAGAAGCGCCCTGGACCAAGAC	983	CATCGAGGGCAAGCAACGCGTGACTCGGGTATCGAAGTGTCTGGACCCCAAA	1046	CCCTACTCANTGAGCAATAACTTTTTTGAAMACCTCTTTGGTTACAGTGGGAGCTTAC	1043	CCCCAGAAATCGATATGGGCTATTTTCGACCTGCTCTCGGCTACAAATTTGGGAAC	1106	CAAAAGTCCAGCTGGAGCTTATCAGTGGAAACCAAAAAGACGGTCCCGGGGTGCAACAT	1103	AAAGAGTCTCTGGCGGCGCCACCACTTGGGAACCGATTGACATCAAAAGGAACCAAGCC	1156	ACCGGATGCATGATCCCAAGCAAGTCGACGCTCCACTTTATGCTACTACGACCTGGC	1163

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2003, 07:56:28 ; Search time 126.323 Seconds
(without alignments)
7819.780 Million cell updates/sec

Title: US-09-884-889-7
Perfect score: 2238
Sequence: 1 agggaaatcacaaacactc.....accggttgatctgaataa 2238

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/Backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2238	100.0	2238	2	US-08-674-887A-7
2	2238	100.0	2238	3	US-08-951-844-7
3	2238	100.0	2238	4	US-09-412-347-7
4	844.4	37.7	2331	1	US-08-418-782-20
5	844.4	37.7	2331	2	US-08-852-219-20
6	844.4	37.7	4403765	3	US-09-103-840A-2
7	844.4	37.7	4411529	3	US-09-103-840A-1
8	832.4	37.2	2235	1	US-08-418-782-1
9	832.4	37.2	2235	1	US-08-228-662-1
10	832.4	37.2	2235	1	US-08-418-782-2
11	721.8	32.3	2221	1	US-08-852-219-1
12	721.8	32.3	2221	1	US-08-418-782-2
13	721.8	32.3	2221	2	US-08-852-219-2
14	721.8	32.3	4795	2	US-08-313-185-45
15	721.8	32.3	4795	3	US-08-082-614A-45
16	716.6	32.0	4794	2	US-08-459-499-8
17	695.2	31.1	2181	4	US-09-328-352-3989
18	672.2	30.0	2262	2	US-08-674-887A-5
19	672.2	30.0	2262	3	US-08-951-844-5
20	672.2	30.0	2262	4	US-09-412-347-5
21	251.8	11.3	620	2	US-08-757-653-145
22	251.8	11.3	620	2	US-08-757-653-149
23	251.8	11.3	620	4	US-08-520-946-149
24	251.8	11.3	620	4	US-08-520-946-145
25	250.2	11.2	620	2	US-08-757-653-143
26	250.2	11.2	620	2	US-08-757-653-146
27	250.2	11.2	620	2	US-08-757-653-147

c	28	250.2	11.2	620	2	US-08-757-653-150
	29	250.2	11.2	620	4	US-08-520-946-143
	30	250.2	11.2	620	4	US-08-520-946-146
	31	250.2	11.2	620	4	US-08-520-946-147
c	32	250.2	11.2	620	4	US-08-520-946-150
	33	248.6	11.1	620	2	US-08-757-653-144
	34	248.6	11.1	620	2	US-08-757-653-148
c	35	248.6	11.1	620	4	US-08-520-946-144
	36	248.6	11.1	620	4	US-08-520-946-148
c	37	157.2	7.0	391	3	US-09-034-205-1
	38	157.2	7.0	391	3	US-08-934-097A-1
	39	157.2	7.0	391	3	US-08-851-588-1
	40	157.2	7.0	391	4	US-09-677-218B-1
	41	157.2	7.0	391	4	US-09-677-192-1
	42	155.6	7.0	391	3	US-09-034-205-3
	43	155.6	7.0	391	3	US-09-034-205-4
	44	155.6	7.0	391	3	US-08-934-097A-3
	45	155.6	7.0	391	3	US-08-934-097A-4

ALIGNMENTS

RESULT 1
US-08-674-887A-7
; Sequence 7, Application US/08674887A
; Patent No. 5939300
; GENERAL INFORMATION:
; APPLICANT: Robertson, Dan E.
; APPLICANT: Sanyal, Indrajit
; APPLICANT: Adhikari, Robert S.
; TITLE OF INVENTION: CATALASES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FASTSEQ for Windows version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/674,887A
; FILING DATE: 03-JUL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09015/002001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2238 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; NAME/KEY: Coding Sequence
; LOCATION: 1...2235
; US-08-674-887A-7

Query Match 100.0%; Score 2238; DB 2; Length 2238;
Best local Similarity 100.0%; Pred. No. 0;
Matches 2238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGAATAACAACTCAGGATCTTCTACGATATACAAACAACTGGCGGAAATGC 60
Db
QY 1 ATGGAATAACAACTCAGGATCTTCTACGATATACAAACAACTGGCGGAAATGC 60
Db
QY 61 CCTTTTACCGAGGTTGCTTAAAGCAAGTGCAGGTGGCGGCAACAAACAGGATGG 120
Db
QY 61 CCTTTTACCGAGGTTGCTTAAAGCAAGTGCAGGTGGCGGCAACAAACAGGATGG 120
Db
QY 121 TGGCCCAACATGCTCAACCTCGGCTTACGCAACATTCATCGCTATCGGACCAAC 180
Db
QY 121 TGGCCCAACATGCTCAACCTCGGCTTACGCAACATTCATCGCTATCGGACCAAC 180
Db
QY 181 GACCGGATTTTGACTATGCGGAGAGTTTAAAGCTAGATCTGGCAGGTTTAAAG 240
Db
QY 181 GACCGGATTTTGACTATGCGGAGAGTTTAAAGCTAGATCTGGCAGGTTTAAAG 240
Db
QY 241 GACCTGCGAGGCTTAATGACAGATTCACAGGACTGGTGGCAGGATTTAGGCTATT 300
Db
QY 241 GACCTGCGAGGCTTAATGACAGATTCACAGGACTGGTGGCAGGATTTAGGCTATT 300
Db
QY 301 GGCCCTCTCTTATACGATGGGTGGCAGGCGCGGCACTTACGCTATCGGTGATGG 360
Db
QY 301 GGCCCTCTCTTATACGATGGGTGGCAGGCGCGGCACTTACGCTATCGGTGATGG 360
Db
QY 361 CGTGGTGGGCTGGCTCCGGCTCACAGCGCTTGGCGCTCTCAATAGCTGGCAGCAAT 420
Db
QY 361 CGTGGTGGGCTGGCTCCGGCTCACAGCGCTTGGCGCTCTCAATAGCTGGCAGCAAT 420
Db
QY 421 GCCAATCTGGATAAGACGCTTGTCTTGGGCCCTCAACAAATAACGCTGCAAA 480
Db
QY 421 GCCAATCTGGATAAGACGCTTGTCTTGGGCCCTCAACAAATAACGCTGCAAA 480
Db
QY 481 ATCTCTGGGCGGATCTAATGATATCTACAGGAACTAGCTCTGGAACATATGGCTTT 540
Db
QY 481 ATCTCTGGGCGGATCTAATGATATCTACAGGAACTAGCTCTGGAACATATGGCTTT 540
Db
QY 541 AAACTTTTGGTTTGCAGTGGCAGCAGATGTATGGAGCTCTGAAGAGATGTATAC 600
Db
QY 541 AAACTTTTGGTTTGCAGTGGCAGCAGATGTATGGAGCTCTGAAGAGATGTATAC 600
Db
QY 601 TGGGAGCAGAAACGATGGTGGGAGCAAGCGCTATGAAGTGAACGAGCTCGAA 660
Db
QY 601 TGGGAGCAGAAACGATGGTGGGAGCAAGCGCTATGAAGTGAACGAGCTCGAA 660
Db
QY 661 AATCCCTGGGAGCGTACAATGGGACTCATCTATGTAAACCCGAGGACCAACGCG 720
Db
QY 661 AATCCCTGGGAGCGTACAATGGGACTCATCTATGTAAACCCGAGGACCAACGCG 720
Db
QY 721 AAGCCAGACCTATCGCTGGCGGTGATATCGTGAGACTTTTGGCCGAATGGCAATG 780
Db
QY 721 AAGCCAGACCTATCGCTGGCGGTGATATCGTGAGACTTTTGGCCGAATGGCAATG 780
Db
QY 781 AATGACGAAGAACCTGGCTCTCATAGGGGTGGACACCTTCGGAACCAACCTATGT 840
Db
QY 781 AATGACGAAGAACCTGGCTCTCATAGGGGTGGACACCTTCGGAACCAACCTATGT 840
Db
QY 841 GCTGCCGATGGGAGAAATATGGGCGGAGAGCTTGGCGGCGAGCTATGAAGAAATG 900
Db
QY 841 GCTGCCGATGGGAGAAATATGGGCGGAGAGCTTGGCGGCGAGCTATGAAGAAATG 900
Db
QY 901 AGCCTGGGTGGAAAAACCTACGCGACCGGACAGCTGGGATACCATCAGGATGGA 960
Db
QY 901 AGCCTGGGTGGAAAAACCTACGCGACCGGACAGCTGGGATACCATCAGGATGGA 960
Db
QY 961 CTAGAAGGCGCTGGACCAAGACCTTACTCAATGGAGCAATTAACCTTTTGAACCTC 1020
Db
QY 961 CTAGAAGGCGCTGGACCAAGACCTTACTCAATGGAGCAATTAACCTTTTGAACCTC 1020
Db
QY 1021 TTTGGTTAGGAGTGGAGCTTACCAAAAGTCCAGCTGGAGCTTATCAGTGGAAACCAAA 1080
Db
QY 1021 TTTGGTTAGGAGTGGAGCTTACCAAAAGTCCAGCTGGAGCTTATCAGTGGAAACCAAA 1080

QY 1081 GACGGTGGCGGGCTGGCACCATTACCGATGCACATGATCCAGCAAGTCCGACGCTCCA 1140
Db
QY 1081 GACGGTGGCGGGCTGGCACCATTACCGATGCACATGATCCAGCAAGTCCGACGCTCCA 1140
Db
QY 1141 TTTATGCTCACTACGGACCTGGCGCTGGCGCTGGCGATGGACCTGATTTACGAAAAATTTCTCGA 1200
Db
QY 1141 TTTATGCTCACTACGGACCTGGCGCTGGCGCTGGCGATGGACCTGATTTACGAAAAATTTCTCGA 1200
Db
QY 1201 CGGTACTATGAAACCCCTGATGATTTGAGATGCTTTGCGGAAAGCATGGTACAAACTG 1260
Db
QY 1201 CGGTACTATGAAACCCCTGATGATTTGAGATGCTTTGCGGAAAGCATGGTACAAACTG 1260
Db
QY 1261 ACACACAGAGATATGGGACCAAGGTGCTACTTGGGACCAAGAGTGCCTCAGGAGAC 1320
Db
QY 1261 ACACACAGAGATATGGGACCAAGGTGCTACTTGGGACCAAGAGTGCCTCAGGAGAC 1320
Db
QY 1321 CTCATCTGGCAAGACCTTATACAGATGTAAAGCATCTCTTGTAGACGAAACGATATT 1380
Db
QY 1321 CTCATCTGGCAAGACCTTATACAGATGTAAAGCATCTCTTGTAGACGAAACGATATT 1380
Db
QY 1381 GAAGGCTTAAAGCCCAAAATCCTGGAATCGGACTGACGTTAAGCGAGCTGGTAAACAGC 1440
Db
QY 1381 GAAGGCTTAAAGCCCAAAATCCTGGAATCGGACTGACGTTAAGCGAGCTGGTAAACAGC 1440
Db
QY 1441 GCATGGCTCTGCACTACTTTTAGAACTCTGACAGCGCGGCGGTGCCAACGTTGCA 1500
Db
QY 1441 GCATGGCTCTGCACTACTTTTAGAACTCTGACAGCGCGGCGGTGCCAACGTTGCA 1500
Db
QY 1501 CGTATACGACTGGCCCCACAAAGACTGGGAAGTAAACACCTTCAGCAACTTGCACAG 1560
Db
QY 1501 CGTATACGACTGGCCCCACAAAGACTGGGAAGTAAACACCTTCAGCAACTTGCACAG 1560
Db
QY 1561 GTACTCAAAACACTAGAGGTATCCAGGAGACTTTAACAGGCGCAATCAGATAACAAA 1620
Db
QY 1561 GTACTCAAAACACTAGAGGTATCCAGGAGACTTTAACAGGCGCAATCAGATAACAAA 1620
Db
QY 1621 GCAGTATCGTTGGCCGACCTGATTTGCTGGCGGCTGTGCGGCTGTAGAAAAAGTGA 1680
Db
QY 1621 GCAGTATCGTTGGCCGACCTGATTTGCTGGCGGCTGTGCGGCTGTAGAAAAAGTGA 1680
Db
QY 1681 AAAGTATCGTGGCCATGAGTGCAGGTGCTTTCAACCCGCGGAGAGCGGATGCCACGCT 1740
Db
QY 1681 AAAGTATCGTGGCCATGAGTGCAGGTGCTTTCAACCCGCGGAGAGCGGATGCCACGCT 1740
Db
QY 1741 GAGCAACCCGATGTGGAAGCTTTGGAAGCTATAGGCGGCTGTGAGACCGGCGGCTT 1800
Db
QY 1741 GAGCAACCCGATGTGGAAGCTTTGGAAGCTATAGGCGGCTGTGAGACCGGCGGCTT 1800
Db
QY 1801 TACATTAACCCGAGCATAAAGTATCCGCTGAGGAAATGCTGCTAGACCGGCGGCTT 1860
Db
QY 1801 TACATTAACCCGAGCATAAAGTATCCGCTGAGGAAATGCTGCTAGACCGGCGGCTT 1860
Db
QY 1861 CTGTCGCTTTGGGACCAAGATGATGCTTTGGTAGGGGCTATGCTGCTAGGCGCACC 1920
Db
QY 1861 CTGTCGCTTTGGGACCAAGATGATGCTTTGGTAGGGGCTATGCTGCTAGGCGCACC 1920
Db
QY 1921 AACTAGCAGGTTTCGACGATGAGTGTACAAATAAGCCGGGTGACGCTATCCAATGAC 1980
Db
QY 1921 AACTAGCAGGTTTCGACGATGAGTGTACAAATAAGCCGGGTGACGCTATCCAATGAC 1980
Db
QY 1981 TTTCTTTTAAACCTGCTAGACCTCAACACTAAATGGGAGCCAGCGATGAATCAGACAA 2040
Db
QY 1981 TTTCTTTTAAACCTGCTAGACCTCAACACTAAATGGGAGCCAGCGATGAATCAGACAA 2040
Db
QY 2041 GTTTTGAAGGCGAGACTTCAAAACTTGGGAGCCAGCGATGAATCAGACAA 2100
Db
QY 2041 GTTTTGAAGGCGAGACTTCAAAACTTGGGAGCCAGCGATGAATCAGACAA 2100
Db
QY 2101 CTGATCTTCGATCCAATTTCCGAGCTTAAGAGCCCTCGCAGAAAGTGTACGGCTGTGACAT 2160
Db
QY 2101 CTGATCTTCGATCCAATTTCCGAGCTTAAGAGCCCTCGCAGAAAGTGTACGGCTGTGACAT 2160
Db
QY 2161 TCTGAGAAAGATTTGTTAAAGATTTTGTGAAGGCTTGGGCCAAAGTAAATGACCTGGAC 2220

Db 2161 TCTGAAGAAAGTTTGTAAAGATTTTGAAGAGCTGGCCCAAGTAATGACCTGGAC 2220
QY 2221 CGGTTTGATCTGAATAA 2238
Db 2221 CGGTTTGATCTGAATAA 2238
RESULT 2
US-08-951-844-7
; Sequence 7, Application US/08951844
; Patent No. 6074860
; GENERAL INFORMATION:
; APPLICANT: Robertson et al.
; TITLE OF INVENTION: Catalogs
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/951,844
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/674,887
; FILING DATE: July 3, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles J. Hellon
; REGISTRATION NUMBER: 28,019
; REFERENCE/DOCKET NUMBER: 331400-55
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2238 NUCLEOTIDES
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: DNA
US-08-951-844-7
Query Match 100.0%; Score 2238; DB 3; Length 2238;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGAATAATCAAAACACTCAGGATCTTCTACGTATTAACAAACACACTGCGCGGAAATGC 60
Db 1 ATGGAATAATCAAAACACTCAGGATCTTCTACGTATTAACAAACACACTGCGCGGAAATGC 60
QY 61 CCTTTTACCGAGGTTTCGCTTAAGCAAGTGCAGGTGCGCGCACCAAAACAGGATTTGG 120
Db 61 CCTTTTACCGAGGTTTCGCTTAAGCAAGTGCAGGTGCGCGCACCAAAACAGGATTTGG 120
QY 121 TGGCCCAACATGCTCAACCTCGGCATCTTACGCCAACATTCATCGCTATCGGACCCAAAC 180
Db 121 TGGCCCAACATGCTCAACCTCGGCATCTTACGCCAACATTCATCGCTATCGGACCCAAAC 180
QY 181 GACCCGGATTTTGACTATCGCGAAGAGTTTAAGAGCTTAGATCTGGCAGCGGTTTAAAG 240
Db 181 GACCCGGATTTTGACTATCGCGAAGAGTTTAAGAGCTTAGATCTGGCAGCGGTTTAAAG 240
QY 241 GACCTGGCAGCGCTTAATGACAGATTCACAGGACTGGTGGCCAGCAGATACGGTCATTAT 300

Db 241 GACCTGGCAGCGCTTAATGACAGATTCACAGGACTGGTGGCCAGCAGATTACGGTCATTAT 300
QY 301 GGGCCCTTCTTTATACGCATGGGTGGCACAGCGCGGCACCTACCGTATCGGTGATGCG 360
Db 301 GGGCCCTTCTTTATACGCATGGGTGGCACAGCGCGGCACCTACCGTATCGGTGATGCG 360
QY 361 CGTGGTGGCGGTGGCTCCGGCTCACAGCGCTTCCGCCCTCTCAATAGCTGGCCAGACAAT 420
Db 361 CGTGGTGGCGGTGGCTCCGGCTCACAGCGCTTCCGCCCTCTCAATAGCTGGCCAGACAAT 420
QY 421 GCCAATCTGGATAAAGCAGCGTTGCTTCTTTGGCCCATCAACAAAAAATACGGTCGAAAA 480
Db 421 GCCAATCTGGATAAAGCAGCGTTGCTTCTTTGGCCCATCAACAAAAAATACGGTCGAAAA 480
QY 481 ATCTCTGGCGGATCTAATGATACACAGGAAACGTAGCTCTGGAACATATGGGCTTT 540
Db 481 ATCTCTGGCGGATCTAATGATACACAGGAAACGTAGCTCTGGAACATATGGGCTTT 540
QY 541 AAAAATTTTGGTTTGCAGGTGGCAGACAGATGTATGGAGGCTGAAGAAGATGTATAC 600
Db 541 AAAAATTTTGGTTTGCAGGTGGCAGACAGATGTATGGAGGCTGAAGAAGATGTATAC 600
QY 601 TGGGAGACAGAAACCGAATGGCTGGGAGACAAGCGCTATGAAGGTGACCGAGAGCTCGAA 660
Db 601 TGGGAGACAGAAACCGAATGGCTGGGAGACAAGCGCTATGAAGGTGACCGAGAGCTCGAA 660
QY 661 AATCCCTTGGAGCGGTACAAATGGGACTCATCTATGTAACCCCGAGACCCACCGGC 720
Db 661 AATCCCTTGGAGCGGTACAAATGGGACTCATCTATGTAACCCCGAGACCCACCGGC 720
QY 721 AAGCCAGACCTATCGCTGCTGGCGTGATATTCGTGAGACTTTTGGCCGAATGGCAATG 780
Db 721 AAGCCAGACCTATCGCTGCTGGCGTGATATTCGTGAGACTTTTGGCCGAATGGCAATG 780
QY 781 AATGACGAAGAAACCGTGGCTCTCATAGCGGGTGGACACACCTTCGGAACAAACCCATGGT 840
Db 781 AATGACGAAGAAACCGTGGCTCTCATAGCGGGTGGACACACCTTCGGAACAAACCCATGGT 840
QY 841 GCTGCCGATGGGAGAAATATGTGGCGGAGAGCGCTGCGCGCGCAGGTATTTGAAGAAATG 900
Db 841 GCTGCCGATGGGAGAAATATGTGGCGGAGAGCGCTGCGCGCGCAGGTATTTGAAGAAATG 900
QY 901 AGCCTGGGTGGAAACACCTACGCGACCGGACAGGTGGGATACCATCACCAGTGA 960
Db 901 AGCCTGGGTGGAAACACCTACGCGACCGGACAGGTGGGATACCATCACCAGTGA 960
QY 961 CTAGAAGCGCTGGACCAAGACCCCTACTCAATGGAGCAATAACTTTTTTGAACACCTC 1020
Db 961 CTAGAAGCGCTGGACCAAGACCCCTACTCAATGGAGCAATAACTTTTTTGAACACCTC 1020
QY 1021 TTTGGTTACGAGTGGGAGCTTACCAAAAGTCCAGCTGGAGCTTATCAGTGGAAACCAAAA 1080
Db 1021 TTTGGTTACGAGTGGGAGCTTACCAAAAGTCCAGCTGGAGCTTATCAGTGGAAACCAAAA 1080
QY 1081 GACGTGCCGGGTGGCACCACATACCGGATGCATGATCCAGCAAGTCCGACCGCTCCA 1140
Db 1081 GACGTGCCGGGTGGCACCACATACCGGATGCATGATCCAGCAAGTCCGACCGCTCCA 1140
QY 1141 TTTATGCTCACTACGAGCTGGCGCTGGCGATGGACCTGTATACGAAATAATTTCTCGA 1200
Db 1141 TTTATGCTCACTACGAGCTGGCGCTGGCGATGGACCTGTATACGAAATAATTTCTCGA 1200
QY 1201 CGGTACTATGAAACCCCTGATGAGTTTCAGATGTTTCGCGAAGAGCATGGTACAAATG 1260
Db 1201 CGGTACTATGAAACCCCTGATGAGTTTCAGATGTTTCGCGAAGAGCATGGTACAAATG 1260
QY 1261 ACACACAGATATGGGACCAAGTGGCTACCTGGGACCAAGAGTGGCTCAGGAAGAC 1320
Db 1261 ACACACAGATATGGGACCAAGTGGCTACCTGGGACCAAGAGTGGCTCAGGAAGAC 1320
QY 1321 CTCATCTGGCAGACCCCTATACAGATGTAAGCCATCTCTTGTAGAGCAAAACCATATT 1380
Db 1321 CTCATCTGGCAGACCCCTATACAGATGTAAGCCATCTCTTGTAGAGCAAAACCATATT 1380

Db 1321 CTGATCTGGCAAGACCTATACAGATGTAAAGCATCTCTGTAGACGAAACGATATT 1380
 QY 1381 GAAGGGCTAAAGCCAAAATCTTGAATCGGACTGACGCTAAGCAGCTGGTAAGCAGC 1440
 Db 1381 GAAGGGCTAAAGCCAAAATCTTGAATCGGACTGACGCTAAGCAGCTGGTAAGCAGC 1440
 QY 1441 GCATGGCTCTGCATCTACTTTAGAACTCTGACAAAGCGGGGCTGCCACGGTGCA 1500
 Db 1441 GCATGGCTCTGCATCTACTTTAGAACTCTGACAAAGCGGGGCTGCCACGGTGCA 1500
 QY 1501 CGTATACGACTGGCCCAACAAAACACTGGGAAGTAAACAAACCTTCAGCAACTTCCAGG 1560
 Db 1501 CGTATACGACTGGCCCAACAAAACACTGGGAAGTAAACAAACCTTCAGCAACTTCCAGG 1560
 QY 1561 GTACTCAAAACACTAGAAAGGTATCCAGGAGACTTTTACCAAGCGGCAATCAGATAACAAA 1620
 Db 1561 GTACTCAAAACACTAGAAAGGTATCCAGGAGACTTTTACCAAGCGGCAATCAGATAACAAA 1620
 QY 1621 GCAGTATCGTTGGCCGACCTGATGTGTGCGCCGCTGTGCGGCTGTAGAAAAGCTGCA 1680
 Db 1621 GCAGTATCGTTGGCCGACCTGATGTGTGCGCCGCTGTGCGGCTGTAGAAAAGCTGCA 1680
 QY 1681 AAAGATGCTGCCATGAGTGCAGGTGCTTTCAGGACACTAGAGCCGCTGACGGCTTTAGAAAC 1740
 Db 1681 AAAGATGCTGCCATGAGTGCAGGTGCTTTCAGGACACTAGAGCCGCTGACGGCTTTAGAAAC 1740
 QY 1741 GACCAACCGATGTGAAGCTTTTGAAGCACTAGAGCCGCTTCAACCCGGGAGCGGATGCCCGCT 1800
 Db 1741 GACCAACCGATGTGAAGCTTTTGAAGCACTAGAGCCGCTTCAACCCGGGAGCGGATGCCCGCT 1800
 QY 1801 TACATTAACCCGAGCATAAAGTATCCGCTGAGGAAATGCTGTAGACCGGCGAGCTT 1860
 Db 1801 TACATTAACCCGAGCATAAAGTATCCGCTGAGGAAATGCTGTAGACCGGCGAGCTT 1860
 QY 1861 CTGTCGCTTTCGGCACCAGAAATGACTGCTTTGGTAGGGGATGCGGTACTGGGCACC 1920
 Db 1861 CTGTCGCTTTCGGCACCAGAAATGACTGCTTTGGTAGGGGATGCGGTACTGGGCACC 1920
 QY 1921 AACTACGAGCTTCCAGCATGGAGTGTGTAAATAAGCCGGTCACTATCCAAATGAC 1980
 Db 1921 AACTACGAGCTTCCAGCATGGAGTGTGTAAATAAGCCGGTCACTATCCAAATGAC 1980
 QY 1981 TTCTTTTGAACCTGTAGACTTCAAACTGGGAACTAAAGTGGGAGTGGCACTCCGGGTAGAC 2040
 Db 1981 TTCTTTTGAACCTGTAGACTTCAAACTGGGAACTAAAGTGGGAGTGGCACTCCGGGTAGAC 2040
 QY 2041 GTTTTGAAGCAGAGACTTCAAACTGGGAACTAAAGTGGGAGTGGCACTCCGGGTAGAC 2100
 Db 2041 GTTTTGAAGCAGAGACTTCAAACTGGGAACTAAAGTGGGAGTGGCACTCCGGGTAGAC 2100
 QY 2101 CTGATCTTCGGATCCAAATCCGAGCTAAGAGCCCTCGCAGAACTGTACGGCTGTGCAGAT 2160
 Db 2101 CTGATCTTCGGATCCAAATCCGAGCTAAGAGCCCTCGCAGAACTGTACGGCTGTGCAGAT 2160
 QY 2161 TCTGAAGAAAAGTTTCTTAAAGATTTTGAAGGCTGGCCCAAGTAATGACCTGGAC 2220
 Db 2161 TCTGAAGAAAAGTTTCTTAAAGATTTTGAAGGCTGGCCCAAGTAATGACCTGGAC 2220
 QY 2221 CGGTTGATCTGAATAA 2238
 Db 2221 CGGTTGATCTGAATAA 2238

RESULT 3

US-09-412-347-7
 ; Sequence 7, Application US/09412347
 ; Patent No. 6410290
 ; GENERAL INFORMATION:
 ; APPLICANT: Robertson, Dan E.
 ; APPLICANT: Sanjay, Indrajit
 ; APPLICANT: Adhikari, Robert S.
 ; TITLE OF INVENTION: CATALASES
 ; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson P.C.
 ; STREET: 4225 Executive Square, Suite 1400
 ; CITY: La Jolla
 ; STATE: CA
 ; COUNTRY: US
 ; ZIP: 92037
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: Windows95
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/412,347
 ; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/674,887
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hallie, Ph.D., Lisa A.
 ; REGISTRATION NUMBER: 38,347
 ; REFERENCE/DOCKET NUMBER: 09015/002001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 619/678-5070
 ; TELEFAX: 619/678-5099
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2238 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; FEATURE:
 ; NAME/KEY: Coding Sequence
 ; LOCATION: 1...2235
 ; US-09-412-347-7

Query Match 100.0%; Score 2238; DB 4; Length 2238;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGAAATATCAAAACACTCAGGATCTTCTACGTATACACAAACACTGGCGGAAATGC	60
Db	1	ATGAAATATCAAAACACTCAGGATCTTCTACGTATACACAAACACTGGCGGAAATGC	60
QY	61	CTTTTACCAGGATTCGCTTAAGCAAGTGCAGGTGGCGGACCAAAACAGGATGG	120
Db	61	CTTTTACCAGGATTCGCTTAAGCAAGTGCAGGTGGCGGACCAAAACAGGATGG	120
QY	121	TGGCCCAACATGCTCAACCTCGGCATCTTACGCCAACATTCATCGGTATCGGACCAAC	180
Db	121	TGGCCCAACATGCTCAACCTCGGCATCTTACGCCAACATTCATCGGTATCGGACCAAC	180
QY	181	GACCCGGATTTTACATATCCGAGATTTAGAGCTAGATCTGGCAGCGGTAAAG	240
Db	181	GACCCGGATTTTACATATCCGAGATTTAGAGCTAGATCTGGCAGCGGTAAAG	240
QY	241	GACCTGGCAGCGCTAATGACATTCACAGACTGGTGGCCAGCAGATTACGGTCATTAT	300
Db	241	GACCTGGCAGCGCTAATGACATTCACAGACTGGTGGCCAGCAGATTACGGTCATTAT	300
QY	301	GGCCCTCTCTTTATACGATGGCTGGCAGCGCGCACCTACCGTATCGGTATGGC	360
Db	301	GGCCCTCTCTTTATACGATGGCTGGCAGCGCGCACCTACCGTATCGGTATGGC	360
QY	361	CGTGTGGCGGTGGCTCCGGCTCAGCGCTTCGCGCTCTCAATAGTGGCCAGCAAT	420
Db	361	CGTGTGGCGGTGGCTCCGGCTCAGCGCTTCGCGCTCTCAATAGTGGCCAGCAAT	420
QY	421	GCCAATCTGGATAAAGCAGCTTCTTCTTGGCCCAATCAAAACAAATACGGTCGAAA	480
Db	421	GCCAATCTGGATAAAGCAGCTTCTTCTTGGCCCAATCAAAACAAATACGGTCGAAA	480

QY 481 ATCTCTGGCGGATCTAATGATACACTACAGGAAACGTAGCTCTGSAACATATGGCTTT 540
 DB 481 ATCTCTGGCGGATCTAATGATACACTACAGGAAACGTAGCTCTGSAACATATGGCTTT 540
 QY 541 AAAAATTTTGGTTTTCAGAGTGGCAGACAGATGATGGAGCCCTGAAGAAGATATATAC 600
 DB 541 AAAAATTTTGGTTTTCAGAGTGGCAGACAGATGATGGAGCCCTGAAGAAGATATATAC 600
 QY 601 TGGGAGCAGAAACCGAATGGCTGGGAGACAAAGCCCTATGAAGGTGACCGAGAGCTCGAA 660
 DB 601 TGGGAGCAGAAACCGAATGGCTGGGAGACAAAGCCCTATGAAGGTGACCGAGAGCTCGAA 660
 QY 661 AATCCCTTGGGAGCGGTCAAAATGAGTCAATCTATGTAAGCCCAAGGACCCCAACGCG 720
 DB 661 AATCCCTTGGGAGCGGTCAAAATGAGTCAATCTATGTAAGCCCAAGGACCCCAACGCG 720
 QY 721 AAGCCAGACCCCTATCGCTGCTGCGGTGATATTCGTGAGACATTTTGGCCGAATGCAATG 780
 DB 721 AAGCCAGACCCCTATCGCTGCTGCGGTGATATTCGTGAGACATTTTGGCCGAATGCAATG 780
 QY 781 AATGACGAAGAAACCGTGGCTCTCATAGCGGTGGACACACCTTCGGAAAAACCCATGGT 840
 DB 781 AATGACGAAGAAACCGTGGCTCTCATAGCGGTGGACACACCTTCGGAAAAACCCATGGT 840
 QY 841 GCTGCGGATGCGGAGAAATATGTCGCGGAGAGCCCTGCGCGCAGGTATTGAAGAAATG 900
 DB 841 GCTGCGGATGCGGAGAAATATGTCGCGGAGAGCCCTGCGCGCAGGTATTGAAGAAATG 900
 QY 901 AGCCTGGGTGGAACCAACACCTACGCGACCGGACACGGTGGCGGATACCATACCAAGTGA 960
 DB 901 AGCCTGGGTGGAACCAACACCTACGCGACCGGACACGGTGGCGGATACCATACCAAGTGA 960
 QY 961 CTAGAAGCGCCTGAGCAACGACCCCTACTCAATGAGCAATPACTTTTGAACCTTC 1020
 DB 961 CTAGAAGCGCCTGAGCAACGACCCCTACTCAATGAGCAATPACTTTTGAACCTTC 1020
 QY 1021 TTTGTTAGAGTGGAGCTTACCAAAAGTCCAGCTGGAGCTTACAGTGGAAACCAAA 1080
 DB 1021 TTTGTTAGAGTGGAGCTTACCAAAAGTCCAGCTGGAGCTTACAGTGGAAACCAAA 1080
 QY 1081 GACGCTGCGGGCTGGCACCATACCGATGACATGATCCAGCAAGTCCGACGCTCCA 1140
 DB 1081 GACGCTGCGGGCTGGCACCATACCGATGACATGATCCAGCAAGTCCGACGCTCCA 1140
 QY 1141 TTTATGCTCACTACGACCTGCGCTGCGCATGGACCTGATACGAAAAAATTTCTCGA 1200
 DB 1141 TTTATGCTCACTACGACCTGCGCTGCGCATGGACCTGATACGAAAAAATTTCTCGA 1200
 QY 1201 CGGTACTATGAACCCCTGATCAGTTTGCAGATGCTTTTCGCGAAAGCATGTACAACTG 1260
 DB 1201 CGGTACTATGAACCCCTGATGAGTTTGCAGATGCTTTTCGCGAAAGCATGTACAACTG 1260
 QY 1261 ACACACAGAGATATGGGACCAAGTGGCTACCTGGGACCAAGATGCTCAGGAAGAC 1320
 DB 1261 ACACACAGAGATATGGGACCAAGTGGCTACCTGGGACCAAGATGCTCAGGAAGAC 1320
 QY 1321 CTCATCTGGCAAGACCTATACAGATGTAAGCATCTCTCTGTAGACGAAAAACGATAT 1380
 DB 1321 CTCATCTGGCAAGACCTATACAGATGTAAGCATCTCTCTGTAGACGAAAAACGATAT 1380
 QY 1381 GAAGGCTTAAAGCCAAATCTGGAATCGGGACTGACGTAAGCGCTGTAGACAG 1440
 DB 1381 GAAGGCTTAAAGCCAAATCTGGAATCGGGACTGACGTAAGCGCTGTAGACAG 1440
 QY 1441 GCATGGCTTCTGCATCTACTTTTGAAGAACTCTGCAAGCGGGGGTGGCAACGCTGCA 1500
 DB 1441 GCATGGCTTCTGCATCTACTTTTGAAGAACTCTGCAAGCGGGGGTGGCAACGCTGCA 1500
 QY 1501 COTATACGACTGGCCCAACAAAGACTGGGAAGTAAACAAACCCCTCAGCAACTTGGCAGG 1560
 DB 1501 COTATACGACTGGCCCAACAAAGACTGGGAAGTAAACAAACCCCTCAGCAACTTGGCAGG 1560
 QY 1561 GTACTCAAAACACTAGAGGTATCCAGGAGGACTTTAACCCAGCGCAATCAGATACAAA 1620

DB 1561 GTACTCAAAACACTAGAGGTATCCAGGAGGACTTTAACCCAGCGCAATCAGATACAAA 1620
 QY 1621 CGAGTATCGTTGCGCGACCTGATTGCTGCGCGGCTGTGCGGGTGTAGAAAAAGCTGCA 1680
 DB 1621 CGAGTATCGTTGCGCGACCTGATTGCTGCGCGGCTGTGCGGGTGTAGAAAAAGCTGCA 1680
 QY 1681 AAGATGCTGGCCATGAGGTGAGGTGCTTTCAACCCGGGACGAGCGGATGCCACGCT 1740
 DB 1681 AAGATGCTGGCCATGAGGTGAGGTGCTTTCAACCCGGGACGAGCGGATGCCACGCT 1740
 QY 1741 GAGCAACCGATGTCGAGCTTTGGAAGCCTTCGAAGCCTTAGAGCCAGCGCTTTAGAAAC 1800
 DB 1741 GAGCAACCGATGTCGAGCTTTGGAAGCCTTAGAGCCAGCGCTTTAGAAAC 1800
 QY 1801 TACATTAACCCGAGCATAAAGTATCCGCTGAGGAAATCTCTGAGACCGGCGCAGCTT 1860
 DB 1801 TACATTAACCCGAGCATAAAGTATCCGCTGAGGAAATCTCTGAGACCGGCGCAGCTT 1860
 QY 1861 CTGCTGCTTTCGGCACCAGAAATGACTGCTTTGGTAGGCGGTATGGGTACTGGGCACC 1920
 DB 1861 CTGCTGCTTTCGGCACCAGAAATGACTGCTTTGGTAGGCGGTATGGGTACTGGGCACC 1920
 QY 1921 AACTAGCAGGTTCCGACGATGAGTGTGTTTACAAATAAGCCGGTCAGCTATCCATGAC 1980
 DB 1921 AACTAGCAGGTTCCGACGATGAGTGTGTTTACAAATAAGCCGGTCAGCTATCCATGAC 1980
 QY 1981 TTTCTTTGTAACCTGCTAGACCTCAACACTAAATGCGAGCCAGCGCATGAATCAGACAAA 2040
 DB 1981 TTTCTTTGTAACCTGCTAGACCTCAACACTAAATGCGAGCCAGCGCATGAATCAGACAAA 2040
 QY 2041 GTTTTGAAGGACGAGACTTCAAACTGCGGAAGTAAAGTGGAGTGGCACCCGGGTAGAC 2100
 DB 2041 GTTTTGAAGGACGAGACTTCAAACTGCGGAAGTAAAGTGGAGTGGCACCCGGGTAGAC 2100
 QY 2101 CTGATCTTCGAGTCCAAATTCGAGCTTAAGCCCTCGCAGAGTGTACGCTGTGCAGAT 2160
 DB 2101 CTGATCTTCGAGTCCAAATTCGAGCTTAAGCCCTCGCAGAGTGTACGCTGTGCAGAT 2160
 QY 2161 TCTGAAGAAAGTTTGTAAAGATTTTGAAGGCTTGGCCCAAGTAATGACCTGGAC 2220
 DB 2161 TCTGAAGAAAGTTTGTAAAGATTTTGAAGGCTTGGCCCAAGTAATGACCTGGAC 2220
 QY 2221 CGGTTGATCTGAATAA 2238
 DB 2221 CGGTTGATCTGAATAA 2238
 RESULT 4
 US-08-418-782-20
 ; Sequence 20, Application US/08418782
 ; Patent No. 5658733
 ; GENERAL INFORMATION:
 ; APPLICANT: Cockerill, Franklin R.
 ; APPLICANT: Kline, Bruce C.
 ; APPLICANT: Uhl, James R.
 ; TITLE OF INVENTION: Detection of Isoniazid Resistant Strains
 ; NUMBER OF SEQUENCES: 22
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Schwegman, Lundberg & Woessner
 ; STREET: 3500 IDS Center
 ; CITY: Minneapolis
 ; STATE: MN
 ; COUNTRY: USA
 ; ZIP: 55402
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/418,782

FILING DATE: 37.78; Score 844.4; DB 1; Length 2331;
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Woessner, Warren D.
 REGISTRATION NUMBER: 30,440
 REFERENCE/DOCKET NUMBER: 150.141US1
 TELEPHONE: 612-339-0331
 TELEFAX: 612-339-0361
 INFORMATION FOR SEQ ID NO: 20:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2331 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 70...2289
 US-08-418-782-20

Query Match
 Best Local Similarity 63.08; Pred. No. 5.4e-251;
 Matches 1390; Conservative 0; Mismatches 801; Indels 15; Gaps 5;
 30 TACGATATACAAACACTGGCGGAAATGCCCTTTTACCGGAGGTTTCGCTTAAGCAAG 89
 93 TACAGAAACACACACCGGAGCCCTAGCAACGGCTGTCCGCTGCTGGGTATATGAATA 152
 90 TGCAGGTGGCGGACCAACAAACAGGATTTGGTGGCCCAACATGCTCAACCTCGGCATCT 149
 153 CCCGTCGAGGCGCGGAGGACAGGACTGGTGGCCCAACCGGCTCAATCTGAAGTACT 212
 150 AGCCCAACATTCATCGCTATCGGACCCCAACGACCCGGATTTGACTATGCCGAGAGTT 209
 213 GCACCAAAACCGGCGCTGCTGAGCCGATGGTGGCGGCTTCGACTATGCCGCGAGGT 272
 210 TAAGAGCTAGATCTGGCAGCGGTATAAAGAGACCTGGCAGCGCTTAATGACAGATTACA 269
 273 CGCGACCATCAGCTTGAGCCCTTGACGCGGACATCGAGGAAGTGATGACCACTCGCA 332
 270 GGACTGTGGCCACAGATTACGCTATTATGGCCCTTCTTTATACGATGGCGTGCA 329
 333 GCGGTGGCGCGGCGGAGTACGCGGCTACGCGGCTGTTTATCCGATGGCGTGCA 392
 330 CAGCGCGCGGACCTACCGTATCGGTGATGCGGCTGGTGGCGGCTTCCGCTCACAGCG 389
 393 CGCTGCGCGACATACCGATCCACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 452
 390 CTTCCGCGCTCTCAATAGCTGGCCAGCAATGCCATCTGGATTAAGACAGCTTGTCT 449
 453 GTTCGCGCGCTTAACAGCTGGCCGACACGCGCTTGGACAGCGCGCGCGCGCGCG 512
 450 TTGGCCCAACAAATAACGTCGAAATAATCTCTGGCGGATCTCAATGATCTAC 509
 513 GTGGCGGCTCAAGAAGAGTACGCGAAGAGCTCTCATGGCGGACCTGATTTTTCG 572
 510 AGGAACCTAGCTCTGGAACATTTGGCTTTAAACATTTTGGCTTTGAGGTGGCAGAGC 569
 573 CGGCAACTGCGGCTGGAATCGATGGCTTCAAGAGCTTCGCGTTCGCTTCGCGCGGT 632
 570 AGATGTATGAGGCTGAAGAAGATGTATCTGGGAGCAGAAACCGAATGGCTGGGAGA 629
 633 CGACCACTGGGAGC---CGATGAGGTCTATTGGGGAAGAGAGCCACCTGGCTCGGCA 689
 630 CAAGCGCTATGAGGTGACCGAGAGCTCGAAATTCCTCGGAGCAGAAACCGAATGGGACT 689
 690 TGAGCGTTACAGCGGTAAGCGGATCTGGAGAACCCGCTGGCGCGGTGAGATGGGCT 749
 690 CATCTATGAAACCCGAGGACCCCAAGCGCAAGCCGATCTATCGCTGCTGCGGTGA 749
 750 GATCTACGTGAACCGGAGGCGGCAACGCGCAACCGCACCCCATGCCGCGCGGTCA 809

QY 750 TATTCCTGACACTTTTGGCCGAATGCGCAATGAATGACGAAGAAACCGTGGCTCTCATAGC 809
 DB 810 CATTCGCGAGACGTTTCGGCGCATGSCCATGACGAGCTGGAAGAACGCGGCTGCTGCT 869
 QY 810 GGCTGACACACTTCGGAAGAACCCATGGTGTGCGGATGCGGAGAAATATGTGGCGCG 869
 DB 870 CGCGGCTACACTTCGGAAGAACCCATGGTGTGCGGATGCGGAGAAATATGTGGCGCG 869
 QY 870 AGAGCCTGCCCGCGGAGTATTGAAGAAATGACCTTGGTTACGAGTGGGAGTACCAAG 1046
 DB 927 CGAACCCGAGGCTGCTCGCTGGAGCAGATGGGCTTGGCTGGAAGAGCTGCTATGGCAC 929
 QY 930 CGGACACGCTGGGATACCATCACAGTGGACTAGAGGCGCTGGACCAAGACCCCTAC 989
 DB 987 CGGAACCGGTAAAGGACGCGATCACAGCGGATCGAGTCTGATGGAGCAACACCCGAC 1046
 QY 990 TCAATGGAGCAATAACTTTTTTGAAGAACTCTTTGGTTACGAGTGGGAGTACCAAG 1049
 DB 1047 GAAATGGAGCAACAGTTTCTCGGATCTCTGAGGCTACGAGTGGAGTGCAGAGAG 1106
 QY 1050 TCCAGCTGGAGCTTATCAGTGGAAACCAAGAGGCTGCGGCGCTGGACCAATACCGGA 1109
 DB 1107 CCTGCTGGCGCTTGGCAATACACCGCAAGGAGCGCGGTGCGGCGACCATCCCGGA 1166
 QY 1110 TGCACATGATCCAGCAAGTGCACGCTTCCATTTATGCTCACTACGAGCTGGCGGTGG 1169
 DB 1167 CCGTTGCG---GCGGCGCAGGCGCTCCCGGAGATGCTGGCAGCTCTGCTGCG 1223
 QY 1170 CATGGACCTGATACGAAATAATTTCTCGAGGCTACTATGAAGAACCTGATGAGTTGCG 1229
 DB 1224 GGTGATCGGCTATGAGCGGATCAGCGTGGCTGGTGAACACCCCGAGAGATTTGGC 1283
 QY 1230 AGATGCTTTCGGAAGCATGTACAACTGACACAGAGATATGGGACCAAGAGTGGC 1289
 DB 1284 CGACAGTTCGCAAGGCTGTGTACAGCTATCCACGAGACATGGTCCGTTGCGAG 1343
 QY 1290 CTACTGGGACCAAGTGCCTCAGGAAGACTCATCTGGCAAGACCTATACCAAGATGT 1349
 DB 1344 ATACTGGGCGCTGGTCCCAAGCAGACCTGCTGTGGCAGAGTCCGCTCCCTGCGGT 1403
 QY 1350 AAGCCTATCTTGTAGACGAAACGATATTAAGAGGCTTAAAGCCAAATCTCTGGAATC 1409
 DB 1404 CAGCAGGACTCTGCGCGGAGCCGAGATTTGCAAGCTTTAAGAGCCAGATCCGCGGATC 1463
 QY 1410 GAGACTCAGCGTAAGCAGCTGTAGCAGGCTGCGGCTTCTGCACTCTACTTTAGAAA 1469
 DB 1464 GGGATGACTGTCTACAGCTAGTTTCGACCGCATGGGCGGCGGTCTGCTTCCGTGG 1523
 QY 1470 CTCTGCAAGCGGCGGCTCCCAAGCTGACATATACGACTGGCGCCCAAGAAAGCTG 1529
 DB 1524 TAGCGACAGCGGCGGCGGCAACGCTGCTGCACTCCGCTGACGCAACAGTCCGGGTG 1583
 QY 1530 GGAAGTAACAAACCTCA---GCACTTCCAGGCTACTCAAAACACTAGAAAGTATCCA 1586
 DB 1584 GGAGGTCAACGACCCGCGGAGGATCTGGCAAGGCTTTCGCACTCCGCTGGAAGAGATCCA 1643
 QY 1587 GGAGACTTTTAACGAGGCTCAATCAGATAAACAACAGTATGTTGGCGGACCTGATTGT 1646
 DB 1644 GGAATCTTCACTCCGCGGCGGCGGAGCATCAAGTGTCTTTCGCGGACCTCGTGT 1703
 QY 1647 GCTGCGGCTGCTGGGCTGTAGAAAAGCTGCAAAAAGATGCTGCGCATGAGGTGCAAGT 1706
 DB 1704 GCTGCGGCTGCTGGGCTGTAGAAAAGCTGCAAAAAGATGCTGCGCATGAGGTGCAAGT 1763
 QY 1707 GCCTTTTCAACCGGAGCGGATGCGGCTGAGCAACCGATGTGGAAGCTTTTCA 1766
 DB 1764 GCCTTTTCAACCGGAGCGGATGCGGCTGAGCAACCGATGTGGAAGCTTTTCA 1823
 QY 1767 AGCACTAGAGCCAGCGCTGACGCTTTAGAAACTATTAACCCAGAGCATTAAGTATC 1826
 DB 1824 CGTGTGGGAGCCCAAGGAGATGCTTCCGAACTACCTCGGAAGAGGCAACCGTTGCC 1883
 QY 1827 CGCTGAGGAATCTCTGAGACCGGCGCAGCTTCTGCTGCTTTCGCGACCAAGAAATGAC 1886

Db 1884 GGCCGAGTACATGCTGCTCGCAAGGCAAGCACTGCTACGCTCAGTGCCTCGAGTAC 1943
Qy 1887 TGCTTTGGTAGCGGTATGCTGCTAGTGGGCAACCACTACGACGCTTCCGACGATGGAGT 1946
Db 1944 GGTGCTGTAGGTGGCTGCGCTCCGCGGCGCAACTACAAAGCGCTTACCGCTGGGGT 2003
Qy 1947 GTTACAAATAAAGCGGGTACGCTATCAATGACTCTCTTTGTAACCTGCTAGACCTCAA 2006
Db 2004 GTTCAACGAGCCCTCCGAGTCACTGACCAACGACTCTCTGTAACCTGCTCGACATGGG 2063
Qy 2007 CACTAAATGGGAGCCAGCGATGAATCAGACAAAGTTTTTGAAGCGAGACTTCARAA 2066
Db 2064 TATCACTGGGAGCCCTCGCAGCAGATGACGGGACCTTACAGGCAAGGATGGC---AG 2120
Qy 2067 TGGCGAAGTAAAGTGGAGTGGCACCCTGGGTAGACCTGATCTTCGGATCCAAATTCGAGCT 2126
Db 2121 TGGCAAGTGAAGTGGACCGCGCAGCGGTGACTGCTTTCGGGTCCAACTCGGAGTT 2180
Qy 2127 AAGAGCCCTCGCAGNAGTGTACGGCTGTGCGAGATCTGAAGAAAGTTTGTAAAGATT 2186
Db 2181 CGGGCGCTTTCGAGGCTTATGCGCGCGATGACGCGAGCGGAAGTTTCGTGCGAGCTT 2240
Qy 2187 TGTGAAGCCCTGGGCAAGTAATGACCTGACCGGTTTGTATCTG 2232
Db 2241 CGTCTGCTGGGCAAGGATGTAACCTCGACAGGTTTCGACGTG 2286

RESULT 5

US-08-852-219-20
; Sequence 20, Application US/08852219
; Patent No. 5922575
; GENERAL INFORMATION:
; APPLICANT: Cockerill, Franklin R.
; APPLICANT: Kline, Bruce C.
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Isoniazid Resistant Strains
; TITLE OF INVENTION: of M. Tuberculosis
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mueiting, Raasch & Gebhardt, P.A.
; STREET: 119 No. 5922575th Fourth Street, Ste. 203
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,219
; FILING DATE: 07-May-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sandberg, Victoria A.
; REGISTRATION NUMBER: 41,287
; REFERENCE/DOCKET NUMBER: 230.00010130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1226
; TELEFAX: 612-305-1228
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2331 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 70...2289
US-08-852-219-20

Query Match 37.7%; Score 844.4; DB 2; Length 2331;
Best Local Similarity 63.0%; Pred. No. 5.4e-251;
Matches 1390; Conservative 0; Mismatches 801; Indels 15; Gaps 5;
Qy 30 TACGTATATACAAACACTGCGGGAATAATGCCCTTTTACCGGAGGTTTCGCTTAAACAAG 89
Db 93 TACAGAAACCAACACCGGAGCGGTAGCAACGGCTGTCCCGTGGTGCATATGAATA 152
Qy 90 TGCAGTGGCGCACCAAAAACAGGATTTGGTGGCCCAACATGCTCAACCTCGGCATCTT 149
Db 153 CCCGTCGAGGGCGGGAACACAGAGCTGTGGGCCCAACCGGCTCACTCTGAAGGTACT 212
Qy 150 AGCCCAACATTATCGCTATCGGACCCAAAGACCCCGGATTTTGTACTATGCGGAGAGTT 209
Db 213 GCACCAAAACCCGGCGCTCGCTGACCCGATGGTGGCGCTTCGACTATGCCGCGGAGT 272
Qy 210 TAAGNAGTATGATCTGGCAGCGGTTTAAANAGACCTGCGAGCGCTTANTGACAGATTCACA 269
Db 273 CGCGACCATCGACGTTGACGCCCTGACGGGACATCGAGGAAGTGATGATCACCATCGCA 332
Qy 270 GGACTGGTGGCAGCAGATTTACGGTCAATATGGCCCTTCTTTATACGATGCGCTGGCA 329
Db 333 GCCGTGGTGGCCCGGACTAGGGCCACTACGGCGCGCTGTTTATCCGATGGCTGGCA 392
Qy 330 CAGCGCCGCGCACTACCGTATCGGTGATGGCGGTGGTGGCGGTGGCTCCGGCTCAGAGG 389
Db 393 CGCTGCGGCACTACCGCATCCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 452
Qy 390 CTTGCGGCTCTCAATAGCTGGCCAGACAAATGCCAATCTGGATAAGACACGCTTGCCTCT 449
Db 453 GTTCGCGCGCTTAAACAGCTGGCCGCAACACAGCTTGGACAGGCGCGCGCTGCT 512
Qy 450 TTGGCCCATCAACAAAATACGGTCAAAAATCTCTGGCGGATCTAATATACATCTAC 509
Db 513 GTGGCGGTCAAGAAAGTACGGCAAGAGCTCTATGGCGCGACCTGATGTTGTTTCG 572
Qy 510 AGAAACGTAGCTCTGAAACTATGGGCTTTTAAAACTTTTGGTTTTCAGGTGGCAGAG 569
Db 573 CGGCAACTCGCGCTGGAATCGATGGCTTCAAGACGTTTCGGGTTTCGGCTTCGGCGGT 632
Qy 570 AGATGTATGGAGCCTGAAGAGATGTACTGGGAGCAGAAACCGAATGCTGGGAGA 529
Db 633 CGACCACTGGGAGCC---CGATGAGGTCTATTGGGCAAGGAAGCACTGCTCGGCA 689
Qy 630 CAAGCGCTATGAAGTGACCGAGAGCTCGAAAATCCCTGGGAGCGCTACAAATGGGACT 689
Db 690 TGAGCGTTACAGCGGTAGCGGGATCTGGAGAACCGCTGGCGCGGTGCATGGGCT 749
Qy 690 CATCTATTAACCCCGAAGGACCAACGGCAAGCCAGACCTATCGCTGCTGCGGTGA 749
Db 750 GATCTACGTGAACCCGAGGGCGGCAACGGCAACCCGACCCCATGGCGCGCGGTGCA 809
Qy 750 TATTCGTGAGACTTTTGGCCGAATGCAATGATGACGAGAAACCGTGGCTCTCATAGC 809
Db 810 CATTCGAGAGCGTTTCGGCGCATGGCCATGAACACCTCGAAGACGCGCGCTGATCT 869
Qy 810 GGGTGGACACACTTTCGGAACCAACCATGCTGCTGCCGATCGGAGAAATATGTGGCGG 869
Db 870 CGGCGTCACTTTTCGGTGAAGCCCATGGCC---CGGCCCGCGGATCTGCTGGGCCC 926
Qy 870 AGAGCTCGCGCGCAGGTATTGAAGAAATAGCCTGGGTTGGAACCAACACCTACGGCAC 929
Db 927 CGAACCCGAGGCTGCTCCGCTGAGCAGATGGCTTGGCTGGAAGAGCTCGTATGGCAC 986
Qy 930 CGGACACGCTGGGATACCATCACCAGTGGACTAGAGGCGCTGAGCAAGACCCCTAC 989
Db 987 CGGAACCGGTAAGGACGGGATCACCAGCGGCATCGAGGTCTGATGACGACACACCCGAC 1046
Qy 990 TCAATGGAGCAATAACTTTTTTGAACCTCTTTGTACGATGGGAGCTTACCAAAAG 1049
Db 1047 GAAATGGGACAAACAGTTTCTCGAGATCCTGTACGGCTACGAGTGGGAGCTGACGAAG 1106

1050 TCCAGCTGGAGCTTATTCAGTGGAAACCAAAAGACGGTGCCTGGGGCTGGACACCATACCGGA 1109
1107 CCTGCTGGCGCTTGGCAATACACCGCAAGAGACGGCGCGGTGCCGCGACCATCCCGGA 1166
1110 TGCACATGATCCAGCAAGTCGACAGCTCCATTTATGCTCACTACGACACCTGGCGCTGG 1169
1167 CCGGTTCG---GGGGCGACGGCGCTCCGACGATGCTGGCCACATGACCTCTCGTTCGG 1223
1170 CATGGACCTGATAGCAAAAATTTCTCGACGGTACTATGAAAACCCCTGATGAGTTGC 1229
1224 GGTGATCCGATCTATGAGCGGATACGGGTGCTGGTGGACACCCCGAGGAATGGC 1283
1230 AGATGCTTTCCGAAAGCATGTTACAACTGACACACAGATATGGGACCAAAAGTGGC 1289
1284 CGACGAGTTTCGCAAGCGCTGTGACAACTGATCCACCGAGACATGGGTCCCGTTCCGAG 1343
1290 CTACCTGGACAGCAAGTGCCTCAGGAACCTCATCTGCGAAGACCTTATACCAATGT 1349
1344 ATACCTTGGCGCTGCTCCCAAGCAGACCTGCTGTGGCAGATCCCGTCCCTCGGT 1403
1350 AAGCCATCTCTTTAGAGCAAAAGATATTGAAGCCCTTAAAGCCAAATCTTGAATC 1409
1404 CAGCCAGACCTCGTGGGGAAGCGGATTCGACGCTTAAAGCCAGATCGGGCATC 1463
1410 GGGACTGACGTAAGCGAGCTGGTAAGCAGCGCATGGGTCTTGCATCTACTTTAGAAA 1469
1464 GGGATGACTGCTCACAGCTAGTTTCGACCCGATGGCGCGGCTGCTGCTCCGTGG 1523
1470 CTCGACAAGCGCGCGGTGCACCGGTGACGTTATACGCTGACGCTGCCCCACAAAAGACTG 1529
1524 TAGCGACAAGCGCGCGGCAACGGTGGTGGCATCCGCTGCGAGCCACAAAGTCGGGTG 1583
1530 GGAAGTAAACACCTCA---GCAACTTCCAGGGTACTCAAAACACTAGAAAGTATCA 1586
1584 GGAGTCAACGACCCGCGGAGTCTGCGAAGTCAATTCGCACCTTGAAGATCCA 1643
1587 GGAGGACTTTAACACAGCGCAATCAGATAACAAAGCATATGCTTGGCGGACCTGATGT 1646
1644 GGAGTCATCACTCCGCGCGCGGGAACATCAAAAGTGTCTTCCGCGACCTCGTCTGT 1703
1647 GCTGGCGGCTGTGGGTGTAGAAAAGCTGCAAAAGATCTTGCCCATAGGTGCGAGT 1706
1704 GCTCGGTGGCTGTCCGCCATAGAGAAGACGAAAGCGGCTGCGCACACATCACGGT 1763
1707 GCCTTTCAACCCGGAGCGGATGCCAGCTGTAGCAACCGATGTGGAGCTTTGGA 1766
1764 GCCCTTCAACCCGGCGCGGATGCTGCGAGGAACAAACCGACGTTGAATCTTTGC 1823
1767 AGCACTAGACGACGCGCTGACGCTTTAGAAACTACATTAACCCGAGCATAAAGTATC 1826
1824 CGTCTGGAGCCCAAGCAGATGGCTTCCGAAACTACCTCGAAAGGGCAACCGTTGCC 1883
1827 CGCTGAGGAATGCTCGTAGCCGGGCGCAGCTTCTGCTGCTTTCGGCACCAAGAAATGAC 1886
1884 GGCCGAGTACATGCTGTCACAAAGGCAACCTGCTTACGCTCAGTCCCTCGAGATGAC 1943
1887 TGCTTTGGTAGCGGTATGGGTACTGGGCACAACTACGACGGTTCGAGCATGGAT 1946
1944 GGTGCTGGTGGTGGGCTGCGGCTCTCGGCGCAAACTACAAGGCTTACCGCTGGCGT 2003
1947 GTTTACAAATAAGCCGGGTGACGCTATCCCAATGACTTCTTTGTAACCTGCTACACCTCAA 2006
2004 GTTACCGAGGCTCCGAGTCACTGACCAACGACTTCTTCGTGAAGCTGCTGCATGGG 2063
2007 CACTAAATGGCAGCGGATGAATCAGACAAAGTTTGTGAAGCAGAGACTTCAAAAC 2066
2064 TATCACTGGAGCCCTCCGACGAGATACGCGGACCTACCAGGGCAAGATGGC---AG 2120
2067 TGGCAAGTAAAGTGGAGTGGACCGGGTAGACCTGATCTTGGATCAATCCAGCT 2126
2121 TGGCAAGTGAAGTGGAGCGGCGCGGTGGACCTGGTCTTCGGGTCCCAACTCGAGTT 2180
2127 AAGAGCCCTCGCAGAAGTGTACGGCTGTGCAGATTCTGAAGAAAGTTTGTAAAGATTT 2186

Db 2181 CGCGGCTTGTTCGAGGCTATGCGCGCGATGACGCCAGCGAAGTTCTGCGAGGACTT 2240
Qy 2187 TGTGAAGGCTGGGCCCAAGTAATGAGCTGGACCGGTTTGTATCTG 2232
Db 2241 CCGTGGCTGGGACCAAGCTGATGAACCTCGACAGGTTGACGCTG 2286

RESULT 6

US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO. 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 37.7%; Score 844.4; DB 3; Length 4403765;
Best Local Similarity 63.0%; Pred. No. 3.2e-249;
Matches 1390; Conservative 0; Mismatches 801; Indels 15; Gaps 5;

Qy 30 TACGTATACCAACAACTGGCGGAAATGCCCCCTTTACCGGAGGTTGCTTAAAGCAAG 89
Db 2153346 TACAGAAACCCACCCGAGCGGTAGCAACGGCTGTCCCGTCTGCTCATATGAATA 2153287
Qy 90 TCGAGGTGGCGCACCAACACAGGATTTGTTGGCCCAACATGCTCAACCTCGGCATCTT 149
Db 2153286 CCGCTGCGAGGCGGGAACACGAGGACTGTGTGGCCCAACCGCTCAATCTGAAAGTACT 2153227
Qy 150 AGCCCAACATTCATCGCTATCGGACCCAAAGACCGGATTTTGAATATGCGGAGACTT 209
Db 2153226 GCACCAAAACCCGCGCTGCTGACCGGATGGGTGGCGCTTCCGCTATGCGCGAGGT 2153167
Qy 210 TGAAGCTAGATCTGGCAGGGTTAAAAGGACCTGGCAGCGCTAATGACAGATTCA 269
Db 2153166 CGCGACCATCGACGTTGACGCCCTGACGCGGAGCATCGAGAACTGATGACCCTCGCA 2153107
Qy 270 GGACTGTGGCCAGCAGATTTACGTCATTTATGGCCCTTCTTTATACGATGCGGTGGCA 329
Db 2153106 GCGGTGTGGCCCGCCGACTACGCCCACTAGGCCCTGTGTTATCCGATGGGTGGCA 2153047
Qy 330 CAGCGCGGCGACCTTACCGTATCGGTGATGCGGTGGTGGCGGTGGCTCCGGCTCACAGG 389
Db 2153046 CGCTGCGGCGACCTACCGCATCCACGACGCGCGCGCGCGCGCGGCGGCGGATGACG 2152987
Qy 390 CTTGCGGCTCTCAATAGCTGGCAGCAATGCCAATCTGGATAAAGCAGCTTGTCTTCT 449
Db 2152986 GTTTCGCGCGCTTAACAGCTGGCCCGACACGACGAGCTTGGACAAAGCGCGCGGTGCT 2152927
Qy 450 TTGGCCCATCAAAATAATACCGTCGAAAATACTCCTGGCGGATCTTAATGATACTCAC 509
Db 2152926 GTGGCGGCTCAAGAAAGTACGGCAAGAGCTCTCATGCGGAGCTGATTTGTTTTCG 2152867
Qy 510 AGGAAACGAGTCTCTGGAATACTATGGCTTTAAACTTTTGTGTTTTCAGAGGTGGCAGGC 569
Db 2152866 CGGCAACTCGCGCTGGAATCGATGCGCTTCAAGAGCTTCGGGTTCGGCTTCGGCGG 2152807

```
QY 570 AGATGCTATGGAGCCTGAGAGATGTATCTACTGGGGAGCAGCAAAACCCGATGGCTGGGAGA 629
Db 2152806 CGACAGTGGGAGCC---CGATAGGCTCTATTGGGGGAAGGAAGCCACCTGGCTCGGCGA 2152750
QY 630 CAAGCGGTATGAAGGTGACCGAGAGCTCGAAATCCCTGGGAGCGGTACAAATGGGACT 689
Db 2152749 TGAGCGTTACAGCGGTAAAGCGGATCTGGAGAACCCGCTGGCCGGGTGCGAGATGGGGCT 2152690
QY 690 CATCTATGTAACCCCGAAGACCCCAACGGCAAGCAGACCTATCGCTGCTGGCGGTGA 749
Db 2152689 GATCTAGCTGAACCCGAGGGGCGGAACGGCAACCCGACCCCATGGCCGGCGGTGGA 2152630
QY 750 TATTCTGTGAGACTTTTGGCCCAATGGCAATGAATGACGAAGAACCCGTGGCTCTCATAGC 809
Db 2152629 CATTGGCGAGAGCTTTGGCGCATGGCCATGAAGCGATCGTGAACAGCGCGCTGATCGT 2152570
QY 810 GGGTGGACACACTTCGGGAAAACCCATGGTGTGCGGATCGCGGAGAAATATGTGGGCG 869
Db 2152569 CGCGGCTCACACTTCGGTAGACCCATGGCG---CGGCCGGCGGATCTGGTCGGCCC 2152513
QY 870 AGAGCTGCCCGCCGACGATATTGAAGAAATGAGCCTGGGGTGGAAAAACACCTACGGCAC 929
Db 2152512 CGAAACCCGAGGCTGCTCGCTGGAGCAGATGGGCTTGGGCTGGAAGAGCTGATGGCAC 2152453
QY 930 CGGACAGGTCGGATACCATCACCAGTGGACTAGAGGCGCCTGGACCAAGACCCCTAC 989
Db 2152452 CGGAACCGGTGAAGACCGCATCACACGCGCATCGAGGTCTATGGAGCAACACCCCGAC 2152393
QY 990 TCAATGGAGCAATAACTTTTTTGAACCTCTTTGGTTTACCAGTGGGAGCTTACCAAAAG 1049
Db 2152392 GAATGGGACAAAGATTCTCTCGAGATCTCTAGCGGTACAGTGGGAGCTGACGAAG 2152333
QY 1050 TCCAGCTGGAGCTTATCAGTGGAAACCAAAAGACGGTGCCTGGGCTGGACCATACCGGA 1109
Db 2152332 CCCTGCTGGCGCTGGCAATACACCGCAAGGCGGCCGCTGCGCGCACCATCCCGGA 2152273
QY 1110 TGCACATGATCCAGCAAGTCGACGCTCCATTTATGCTACTACGGACCTGGGCTGCG 1169
Db 2152272 CCCGTTGC---GCGGGCGAGGGCGCTCCCGACGATGCTGCCACTGACCTCTCGCTGCG 2152216
QY 1170 CATGACCCCTGATTACGAAAAAATTTCTCGACGGTACTATGAAAAACCTGATAGTTTGC 1229
Db 2152215 GGTGATCCGATCTATGAGCGGATCACGCTGCTGCTGGTGAACACCCCGAGGATTTGC 2152156
QY 1230 AGATGCTTTCCGAAAGCATGGTACAAACTGACACAGAGATATGGACCAAAAGTTCG 1289
Db 2152155 CGACGAGTTCCCAAGGCTGTACAAAGCTGATCCACCGAGACATGGGTCCCGTTGCGAG 2152096
QY 1290 CTACCTGGACAGAGTGCCTCAGGAGACCTCATCTGGCAAGACCTTATACCATGAT 1349
Db 2152095 ATACCTGGGCGGCTGGTCCCAAGCAGACCTGCTGTGGCAGGATCCGGTCCCTGCGGT 2152036
QY 1350 AAGCATCTCTGTAGACGAAACGATATTGAAGGCTTAAAGCCAAATCCCTGGAATC 1409
Db 2152035 CAGCCACGACCTCTGCGGAAGCGGAGATTGCCAGCTTAAGACCGAGATCCGGGCATC 2151976
QY 1410 GGGACTGACGTAAGCAGCTGGTAAGCACCGCATGGGCTTCTGCATCTACTTTTAAAA 1469
Db 2151975 GGGATTACTCTCACAGTAGTTTCGACCGCATGGCGGCGGCTGCTGCTCCGTGG 2151916
QY 1470 CTCTGACAGCGCGGCTGCCAAGGTCGACGATATACGATGGCCGCCCAAAAAGACTG 1529
Db 2151915 TAGCGACAAGCGCGCGCGCAAGGTTGGTTCGATCCGCTGCAAGCCCAAGTTCGGGTG 2151856
QY 1530 GGAAGTAAACACCTCA---GCAACTTGCAGGGTACTCAAAACACTAGAGGTATCCA 1586
Db 2151855 GGAGGTCAACGACCCGACGGGATCTCGCGAAGGTCTATCGACCCCTGGAGAGATCCA 2151796
QY 1587 GGAGACTTTAACCAAGCGCAATCAGATTAACAAAGCAGTATCGTTGGCCGACCTGATTGT 1646
Db 2151795 GGAGTCAATCAACTCCGCGCGCGGGAACATCAAAAGTGTCTTCGCGCGACCTCGTGT 2151736
```

```
QY 1647 GCTGGCGGCTGTGGGGTGTAGAAAAAGCTGCAAAAAGATGCTGGCCATGAGGTGCAGGT 1706
Db 2151735 GCTCGGTGGCTGTCCGCCATAGAGAAAAGCAGAAAGCGGCTGGCCACAACTCACGCT 2151676
QY 1707 GCCTTTTCAACCCGGGACGAGCGGATGCCACCGCTGAGCAAAACCGATGTGGAGCTTTTCCA 1766
Db 2151675 GCCCTTCAACCCGGGCGCCGAGGATGCTGCGAGGAACAACCGAGTGGATCTCTTGGC 2151616
QY 1767 AGCACTAGAGACGCGGCTGACGGCTTTAGAAACTTACATTAAACCGGAGCATAAAGTATC 1826
Db 2151615 CGTGTGGAGCCCAAGCAGATGGCTTCCGAAACTTACCTCGGAAAGGCAACCCGTTGCC 2151556
QY 1827 CGCTGAGGAATATGCTCTAGACCGGGCGCAGCTTCTGCTGCTTGGGACCAAGAAATGAC 1886
Db 2151555 GGCCGAGTACATGTGCTGTCGAAAGGCAACCTGCTTACGCTAGTGGCCCTGAGATGAC 2151496
QY 1887 TGCTTTGGTAGCGGTATGCGGTACTGGGCAACCACTACGACGGTTTCGACGATGGAGT 1946
Db 2151495 GGTGCTGGTAGTGGCTTGGCGTCTCGGCGCAACTTACAAAGCCGCTTACCGCTGGGCT 2151436
QY 1947 GTTTACAAATAAAGCGGCTGAGCTATCAATGACTTTCTTTTAAACCTGCTAGACCTCAA 2006
Db 2151435 GTTCACCGAGGCTCCGAGTCACTGACCAACGACTTCTTCGTGAACCTGCTCGACATGG 2151376
QY 2007 CACTAAATGGCGAGCCAGCGGATGAATCAGACAAAGTTTTTTGAAGCCAGAGACTTCAAAC 2066
Db 2151375 TATCACCTGGGAGCCCTCGCCAGCAGATGACGGGACCTACCAAGGCAAGATGGC---AG 2151319
QY 2067 TGGCAAGCTAAAGTGGAGTGCACCGGCTAGACCTGATCTTCGGATCCAATTCGAGCT 2126
Db 2151318 TGGCAAGTGAAGTGGAGCCGCGCGGCTGGACCTGCTTCGGGTCCAACCTCGAGGT 2151259
QY 2127 AAGAGCCCTCCGAGAAGTGTACGGTGTGCGAGATTCTGAAGAAAAATTTGTTAAAGATTT 2186
Db 2151258 GCGGCGCTTGTGAGGCTTATGGCGCGCATGACGCGCAGCGGAAGTTTCGTGCAAGACT 2151199
QY 2187 TGTGAAGCCCTGGGCCAAAGTAATGGACCTGACCGGTTTGTATCTG 2232
Db 2151198 CGTGGCTGCTGGGCAAGGTGATGAACCTCGACAGTTTCGACGTG 2151153
```

RESULT 7

```
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103, 840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37/RV
US-09-103-840A-1
```

```
Query Match 37.7%; Score 844.4; DB 3; Length 4411529;
Best Local Similarity 63.0%; Pred. No. 3.2e-249;
Matches 1390; Conservative 0; Mismatches 801; Indels 15; Gaps 5;
```

```
QY 30 TACGTATTAACACAAACTGGCGGAAATGCCCTTTTACCAGGAGTTCGCTTAAGCAAG 89
Db 2156086 TACAGAAACCAACCAACCGGCGCTAGCAACGGCTGTCCCGTCTGGGTGATATGAATA 2156027
QY 90 TGCAGGTGGGCGCACCAAAAACAGGGATTGGTGGCCCCAACATGCTCAACCTCGGCATCTT 149
```


Sequence 1, Application US/08418782
 Patent No. 5658733
 GENERAL INFORMATION:
 APPLICANT: Cockerill, Franklin R.
 APPLICANT: Kline, Bruce C.
 APPLICANT: Ohl, James R.
 TITLE OF INVENTION: Detection of Isoniazid Resistant Strains
 TITLE OF INVENTION: of M. Tuberculosis
 NUMBER OF SEQUENCES: 22
 CORRESPONDENCE ADDRESS:
 ADDRESS: Schwegman, Lundberg & Woessner
 STREET: 3500 IDS Center
 CITY: Minneapolis
 STATE: MN
 COUNTRY: USA
 ZIP: 55402

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/418,782
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Woessner, Warren D.
 REGISTRATION NUMBER: 30,440
 REFERENCE/DOCKET NUMBER: 150.141US1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 612-339-0331
 TELEFAX: 612-339-3061

INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2235 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA
 US-08-418-782-1

Query Match 37.2%; Score 832.4; DB 1; Length 2235;
 Best Local Similarity 63.0%; Pred. No. 2,7e-247;
 Matches 1389; Conservative 0; Mismatches 801; Indels 16; Gaps 6;

30 TACGTTATACAAACACTGGCGGAAATGCCCTTTTACCGGAGGTTCGCTTAAGCAAG 89
 33 TACAGAAACCAACCGCGAGCGCTAGCAACGGCTGTCCCGTCTGGTGCATATGAANTA 92
 90 TGCAGTGGCGGCCAACAAACAGGGATTGGTGGCCCAACATGCTCAACCTCGGCATCTT 149
 93 CCCGTCGAGGGCGCGGAAACACAGGACTGGTGGCCCAACCGGCTCAATCTGAAGGTACT 152
 150 ACGCCCAACATTCATCGCTATCGGACCCAAACGACCGGATTTTGACTATGCGGAGAGTT 209
 153 GCACCAAAACCGGCGCGTCTGCTACCCGATGGTGGCGGCTTCGACTATGCGCGGAGGT 212
 210 TAAGAAGCTAGATCTGGCAGCGGTTAAAGAGACCTGGCAGCGCTTAATGACAGATTACA 269
 213 CGCAGCATTCAGCTTGACGCCCTGACCGGGACATCGAGGAAGTGATGACCACCTCGCA 272
 270 GGACTGTGGCAGAGATAGGTCATATGGCCCTTCCTTTATACGATGGCGTGGCA 329
 273 GCGGTGTGGCGCGGACTACGGCCACTACGGCGGCTGTATCCGGATGGCGTGGCA 332
 330 CAGCCCGGCACTACCGTATCGGTGATGGCGGTGGTGGCGGTGGCTCGGCTCACAGG 389
 333 CGCTCCGGCACCTACCGCATCCACGACGGCGCGCGCGCGCGCGGGCGGATGACGG 392
 390 CTTTCGGCGCTCAATAGCTGGCCAGCAAAATGCCAATCTGGATAAAGCAGCGCTTCTTCT 449
 393 GTTCGGCGCGCTTACAGCTGGCCGACACGCCAGCTTGGACAAAGGCGCGCGCTGCT 452

450 TTGGCCCATCAACAAAAATACGGTTCGAAAAATCTCCTGGGGCGGATCTAATGATACTAC 509
 453 GTGGCCGCTCAAGAAGAAGTACGGCAAGAAGCTCTCATGGGCGGACCTGATTTGTTTCG 512
 510 AGGAAACGTAGCTCTGGAAACCTATGGGCTTTTAAAACTTTTGGTTTTCGAGGTGGCAG 569
 513 CGGCAACTGCGCGCTGGAAATCGATGGGCTTCAGACGTTTCGGGTTTCGGCTTCGGCG 572
 570 AGATGTATGGAGCCTGAAGAAGATGTATCTTGGGAGCAGAAACCGAATCTGGGAGA 629
 573 CGACCACTGGGAGCC---CGATGAGTCTATTGGGCAAGGAAGCCACCTGGCTCGGCA 629
 630 CAAGCGCTATGAGGTGACCGAGAGCTCGAAATCCCTGGGAGCGGTACAAATGGGACT 689
 630 TGAGCGTTACAGCGGTAAAGCGGATCTGGAGAACCGCTGGCGCGGTGCAGATGGGGT 689
 690 CATCTATGTAAACCCCGAAGGACCAACGGAAGCAGACCTATCTGCTGCTCGCGTCA 749
 690 GNTCTACGTGAACCGGAGGCGCCACCGCAACCGGACCCCTGCGCGCGCGGTGCA 749
 750 TATTCGTGAGACTTTTGGCCGAATGGCAATGAATGACGAAGAACCGTGGCTCTCATAG 809
 750 CATTCGCGAGACGTTTTCGGCGCATGCCATGAACGAGCTCGAAACACAGCGCGCTGAT 809
 810 GGGTGGACACACCTTCGGGAAACCCATGCTGCTGCCGATGCGGAGAAATATGTGGCG 869
 810 CGGCGGTACACTTTTCGGTAAGACCCATGGCG---CGGCGCGCGCGATCTGCTCGCC 866
 870 AGAGCTGCGCGCGAGGTATTGAAGAATGAGCTGGGTCGGAACACACCTACGGCAC 929
 867 CGAACCCGAGGCTGCTCCGCTGGAGCAGATGGCTTGGCTGGAGAGCTCGTATGGCAC 926
 930 CGGACACGCTGGGATACCATCAGCTGGACTAGAAGCGGCTGACCAAGACCCCTAC 989
 927 CGGAACCGGTAAAGCGGATCACCAGCGGATCGAGGTCGTATGACGAACACCCGAC 986
 990 TCAATGGAGCAATACTTTTGAACACCTCTTTGGTTAGGATGGGAGCTTACCAAG 1049
 987 GAAATGGGACAAACAGTTTCTCGAGATCTCTACGGCTACGATGGGAGCTGACGAAG 1046
 1050 TCCAGCTGGAGCTTATCAGTGGAAACCAAGACGCTGCGGGCTGGCACCATACCGGA 1109
 1047 CCCTGCTGGCTTGGCAATACCGCCAGGACGCGCGGCTGCGGACCATACCCGGA 1106
 1110 TGCACATGATCCAGCAAGTCGACGCTCCATTTATGCTACTAGCAGCTGGCGTGG 1169
 1107 CCGGTTTCG---CGGCGCCAGGCGCTCCCGCAGCATGCTGGCAGCTCTCGGTGG 1163
 1170 CATGACCTGATTCAGGAAATAATTTCTCGAGGTTACTATGAACCCCTGATGAGTTGC 1229
 1164 GGTGATCCGATCTATGAGCGGATCAGCGTGGCTGGGCTGGAACACCCCGAGGAATGG 1223
 1230 AGATGCTTTCGGAAGCATGTGACAACTGACACAGAGATATGGGACCAAGTGG 1289
 1224 CGACGAGTTGCGCAAGGCTGGTACAGCTGATCCACCGAGACATGGGTCGCGTCCGAG 1283
 1290 CTACCTGGGACAGAGTCCCTCAGGAAGCTCATCTCGCAAGACCTTATACCAATGT 1349
 1284 ATACCTTGGCGCTGGTCCCAAGCAGACCTGCTGTGCGAGGATCCCGTCCCTCGGT 1343
 1350 AAGCAATCTCTTAGAGCAAAACGATATTGAAGGCTTAAAGCCCAAAATCTTGAATC 1409
 1344 CAGCCACGACCTCGTGGGGAAGCGGATTTGCCAGCTTAAAGCGCAGATCCGGGCATC 1403
 1410 GGGACTGACGCTAAGCGAGCTGTTAGCAGCGCATGGGCTTCTGCACTCTTATTAGAA 1469
 1404 GGGATGAGTGTCTCACAGCTAGTTTCGACCCGATGGGCGGCGGTCTGCTTCGCTGG 1463
 1470 CTCTCAAGCGCGCGGTGCCAACGCTGACGATGACCTTACGCTGGCCCAACAAACACTG 1529
 1464 TAGCCACAGCGGCGGCGCCACCGGTGCTCGCATCCGCTGCGACCCACAGTCCGGTG 1523
 1530 GGAAGTAAACAACCTCA---GCAACTTGCAGGGGTACTCAAAACACTAGAGGTATCCA 1586


```

Db 1524 GGAGGTCACAGCCCGGACGGGATCGGCAGGTCAATCGCACCTGGAGAGATCCA 1583
QY 1587 GGAGGACTTAAACAGCGCAATCAGATAACAAGCAGTATCGTTGGCGGACCTGATGTT 1646
Db 1584 GGAGTCATTCACTCGCGCGCGCGGGGAACATCAAAGTGTCTTCCGCGGACCTCGTGT 1643
QY 1647 GCTGGCGGCTGTGGGGGTGTAGAAAAGCTGCAAAAAGATGCTGGCGCATGAGTGCAGGT 1706
Db 1644 GCTCGGTGCTGTGGCCATAGAGAAACGCAAGAGGGGTGGCCACAACATCACGTT 1703
QY 1707 GCTTTTCAACCGGACGAGCGGATGCCACCGCTGAGCAAAACGATGTGGAAGCTTTTCA 1766
Db 1704 GCGCTTCACCGCGCGCGCGGATGCGTGCAGGAACAAACCGGAGTGAATCCTTTGC 1763
QY 1767 AGCACTAGAGCAGCGGCTGACGCTTTAGAACTACATTAACCGGACCAATAAGTATC 1826
Db 1764 CTGCTGGAGCCAGGAGATGGCTTCGAAACTACCTCGGAAGGGCAACCCGTTGCC 1823
QY 1827 CGCTGAGGAATGCTCTAGACCGGCGGAGCTTCTGCTGCTTTCGGCACCAAGAAATGAC 1886
Db 1824 GCGCAGTACATGCTGCTCGACAAGCGGAACCTGTAGCTCAGTGCCTCGCTGAGATGAC 1883
QY 1887 TGTCTTGTAGCGGTATGCGTGTACTGGGCACCAACTACGACGTTCCGACATGGAGT 1946
Db 1884 GTGCTGTAGTGGCTGCGCGTCTCGG-GCAAACTACAAGCGCTTACCGCTGGCGT 1942
QY 1947 GTTTACAAATAAGCGGCTCAGCTATCCAATGACTTCTTTGTAACCTGCTAGACCTCA 2006
Db 1943 GTTACCGAGGCTCCGAGTCACTGACCAACGACTTCTCGTGAACCTGCTCGACATGG 2002
QY 2007 CACTAATGCGGAGCGGATGAATCAGACAAAGTTTTTGAAGGACAGAGACTTCMAAAC 2066
Db 2003 TATCACCTGGAGGCGCTCGCCAGCAGATGACGGGACCTACCAGGCAAGGATGGC--AG 2059
QY 2067 TGGCGAAGTAAAGTGGAGTGGCCAGCGGTAGACCTGATCTTCGGATCCAAATCCGAGCT 2126
Db 2060 TGGCAAGTGAAGTGGAGCGCGGACCGGTGAGCTGCTCTCGGTCCTCAACTCGGAGTT 2119
QY 2127 AAGAGCCCTCGCAGAAGTGTACGCTGTGCAAGTCTGAAGAAAGTTTGTAAAGATT 2186
Db 2120 GCGGCGCTTGTGAGTCTATGCGCGGATGACGCGCAGCGGAAGTCTGTCGAGGACTT 2179
QY 2187 TGTGAAGGCTGGCCAAAGTAAATGAGCTGGACCGCGGTTGATCTG 2232
Db 2180 CGTGCCTGGGACAAAGTGTATGAACCTCGACAGGTTCGACGTG 2225

```

RESULT 9

```

US-08-228-662-1
; Sequence 1, Application US/08228662
; Patent No. 5688639
; GENERAL INFORMATION:
; APPLICANT: COCKERILL, FRANKLIN R.
; APPLICANT: KLINE, BRUCE C.
; APPLICANT: UHL, JAMES R.
; TITLE OF INVENTION: DETECTION OF ISONTAZID RESISTANT STRAINS
; TITLE OF INVENTION: OF M. TUBERCULOSIS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCHWEGMAN, LUNDBERG & WOESSNER, P.A.
; STREET: 3500 IDS CENTER
; CITY: MINNEAPOLIS
; STATE: MINNESOTA
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/228,662

```

```

; FILING DATE: 18-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WOESSNER, WARREN D.
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 150.123US01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2235 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-228-662-1

```

```

Query Match 37.2%; Score 832.4; DB 1; Length 2235;
Best Local Similarity 63.0%; Pred. No. 2.7e-247;
Matches 1389; Conservative 0; Mismatches 801; Indels 16; Gaps 6;

```

```

QY 30 TACGTATACACAAACACTGGCGGAAATGCCCTTTTACCGGAGGTTCCGTTAAGCAAG 89
Db 33 TACAGAAACACACACCGGAGCGCTAGCAACGGCTGTCCCGTCTCATATGAAATA 92
QY 90 TGCAGGTGGCGCACAAAACAGGATTTGGTGGCCCAACATGCTCAACCTCGGCATCTT 149
Db 93 CCCCCTCGAGGCGGCGGAAACAGGACTGTGTGGCCCAACCGGCTCAATCTGAAGTACT 152
QY 150 AGCCCAACATTCATCGCTATCGGACCCAAAGACCCGGATTTTGTACTATGCCGAAGATT 209
Db 153 GCACAAACCCGCGCGCTCGCTGACCCGATGGTGGCGGCTTGGACTATGCCGCGAGGT 212
QY 210 TAAGAAGCTAGATCTGGCAGCGGTTTAAAGAGACCTGGCAGGCTAATGACAGATTACA 269
Db 213 CGCAGACCATCAGCTGTGACGCCCTGACCGGGACATCGAGGAAGTATGACCACTCGCA 272
QY 270 GGACTGTGGCCAGCAGATTACGGTCAATATATGGCCCTTCTTTATACGATGGCGTGA 329
Db 273 GCGCTGTGGTGGCGCGGACTACGGCCACTACGGCGCGCTGTTATCCGGATGGCGTGA 332
QY 330 CAGCGCGGACCTACCGTATCGGTATGCGCGGTGGTGGCGGTGGCTCCGGCTCACAGCG 389
Db 333 CGCTGCGCGACCTACCGCATCCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCATCGACG 392
QY 390 CTTCCGCGCTCTCAATAGCTGGCCAGACAATGCCAATCTGGATAAGACACCTTGTCTCT 449
Db 393 GTTCGCGCGCTTAAACAGCTGGCCCGACACAGCCAGCTTGGACANGCGCGCGCGCTGCT 452
QY 450 TTGCCCATCAACAAATAATACGGTTCGAAATACTCCTGGCGGATCTAATGATACTCAC 509
Db 453 GTGCGCGGTCAAGAAGAGTACGGCAAGAGCTCATATGGCGGACCTGATTGTTTCGC 512
QY 510 AGGAACGTAGCTCGAAACTATGGGCTTTAAACCTTTTGGTTTTCAGGTGGCAGAGC 569
Db 513 CGGCAACTGCGGCTGGAATCGATGGGCTTCAAGACGTTTCGGGTTTCGGCTTCGGCGGGT 572
QY 570 AGATGTATGGGAGCCTCAAGAAGATGTATCTGGGAGCAGAACCAACCAATGGCTGGGAGA 629
Db 573 CGACCACTGGAGGCC---CGATGAGTCTATTGGGCAAGNAGCCACCTGGGCTGGCGA 629
QY 630 CAAGCGCTGAAGGTGACCGGAGCTCGAAATATCCCTGGGAGCCCTACAAATGGGACT 689
Db 630 TGAGCGTTACAGCGGTAAAGCGGATCTGGAGAACCCGCTGCGCGGTGCGAGATGGGCT 689
QY 690 CATCTATGTAACCCCGAAGGACCCAAACGCAAGCCAGACCCCTATCGTCTCGCGGTGA 749
Db 690 GATCTAGTGAACCCCGAGGGGCGGCAACCGCAACCCGAGCCCATGCCCGCGGTCGA 749
QY 750 TATTCTGTGACCTTTTGGCGGAATGGAATGATGACGAAGAAACCGTGGCTCTCATAGC 809
Db 750 CATTCCGAGAGCGTTTCGGCGCATGGCCATGAACGAGCTCGAANACAGCGCGCTGATCGT 809

```


Qy	810	GGGTGGACACACCTTCGGAAAAACCCCATGGTGTCTGCCGATCCGAGAGAAATATGTGGGCGG	869
Db	810	CGCGGGTCACACTTTCGGTAAGAACCCATGCGCGC--CGGCCCGCGCGATCTGGTGTGCGCC	866
Qy	870	AGAGCTGCCCGCGCAGGTATTCAAGAATAAGACCTGGGTGGATGAAAAACACCTACGSCAC	929
Db	867	CGAACCGGAGCTCTCCGCTGGAGCAGATGGGCTTGGGCTGGAACAGCTCTGTATGGCAC	926
Qy	930	CGGACACGGTGCGGATACCATCACCACTGGAGTAGAAGCGCCTGGACCAAGACCCTAC	989
Db	927	CGGAACCGGTAAAGACCGGATACCAAGCGGCATCGAGGTCGTATGGACGAACACCCCGAC	986
Qy	990	TCAATGGAGCAATAACTTTTTTCAAACCTCTTTGGTTACGAGTGGGAGCTTACCAAAAG	1049
Db	987	GAATGGGACACAGTTTCTTCAGATCTCTGTACGGCTACGAGTGGGAGCTGACGAAGAG	1046
Qy	1050	TCCAGCTGGAGCTTATTCAGTGGAAACCAAAGACGGTCCCGGGCTTGGACCAATACCGGA	1109
Db	1047	CCCTGCTGGCGCTTGGCAATACACCGCAAGACGCGCGGTGCGGACCAATCCCGA	1106
Qy	1110	TGCACATGATCCAGCAAGTGCAGCGTCCATTTATGTCTACTACGGAAGCTGGCGTGG	1169
Db	1107	CCGCTTGC---GGGGCCAGGCGCTCCCGAGCATGCTGGCCACTGACCTCTCGCTGG	1163
Qy	1170	CATGGACCTGTATTACGAAAAAATTTCTCGACGGTACTATGAAACCCCTGATGAGTTTC	1229
Db	1164	GGTGGATCCGATCTATGAGCGGATACGGGTGCTGGCTGGAAACACCCCGAGGAATTGC	1223
Qy	1230	AGATGCTTTCCGAAAGCATGTTACAACTTGACACACACAGATATGGGACCAAGTGG	1289
Db	1224	CGACGAGTTCCGCCAAGCCCTGGTACAAGTGATCCACCGAGACATGGTCCGTTGGAG	1283
Qy	1290	CTACTGGGACACAGATGGCTCAGGAAGACCTCTCTGGCAAGACCCATATACCGATGT	1349
Db	1284	ATACTTGGGCCGTGTGTCGCCAAGCAGACCCCTGCTGTGGCAGGATCCGGTCCCTGGGT	1343
Qy	1350	AAGCATCTCTTTAGACGAAAAACCATTTGAAGCCCTAAAGCCATAAAGCCAAATCCTGAAATC	1409
Db	1344	CAGCAGGACCTCTGCGGCAAGCCGAGATTGCCGCTTAAGAGCGAGATCCGGGCATC	1403
Qy	1410	GGGACTGACGGTAAGCGAGTGGTTAAGCAGCGCATGGGCTCTTGCACTCTACTTTTAGAAA	1469
Db	1404	GGGATTGACTGTCTACAGCTAGTTTCGACCGCATGGCGCGCGGTCTGCTTCGTCCGTGG	1463
Qy	1470	CTCTGACAGCGGGGGGTGCCAAGCGTGACGTATACGACTGGCCGCCACAAAAGACTG	1529
Db	1464	TAGCGACAAGCGGGGGCGCCAAACGGTGTGTCATCCGCTCCGCTCGACGCAAGTCGGGTG	1523
Qy	1530	GGAAGTAAACAACCCCTCA---GCACTTTGCCAGGTTACTCAAAACACTAGAAGTATCCA	1586
Db	1524	GGAGTCAACGACCCGACGGGGATCTGCCAAGGTCTTCGCACCCTCGAAGAGATCCA	1583
Qy	1587	GGAGGACTTTAAACAGGCGCAATCAGATAACAAGCAGTATCCTTTGGCCGACCTGATTGT	1646
Db	1584	GGAGTCTATTCACTCCGCGCGCGCGGGAACATCAAGTGTCTCTCGCCGACCTCGTCTGT	1643
Qy	1647	GCTGGCGGCTGTGCGGGTGTAGAAAAGCTGCAAAAGATGCTTGGCCATGAGGTGCAGGT	1706
Db	1644	GCTCGTGTGCTGTGCCGCCATAGAGAAGCAGAAAAGCGGGCTTGGCCACAACATCACGCT	1703
Qy	1707	GCTTTTCAACCCGGGAGGAGCGGATGCCACCGCTGAGCAAAACCGCATGTGGAAGCTTTCGA	1766
Db	1704	GCCCTTCAACCCGGGCGGACGAGTGGTTCGAGGAACAAACCGAGCTGGATTCCTTTGC	1763
Qy	1767	AGCACTAGACCCAGCGCTCACGGCTTTAGAAACTACATTTAAACCGGACGATAAAGTATC	1826
Db	1764	CGTGTGGAGCCCAAGCAGATGGCTTCCGAAACTACCTCGGAAAGGGCAACCCCTTGC	1823
Qy	1827	CGCTGAGGAATGCTGTAGACCGGGGCGAGCTTCTGTGCTTTCGGCACCAAGAAATGAC	1886
Db	1824	GGCGGATGATCTGTCTCAACAAGGCAACCTGCTTACGCTCAGTGGCCCTCAGATGAC	1883

Qy	1887	TGCTTTTGGTAGGCGGTATGGGTGTACTTGGGSCACCAACTAGGACGGTTCCGACGATGGAGT	1946
Db	1884	GGTCTGGTAGGTGGCTGCGGCTCTCGG - GCAAACTACAAAGCGCTTACCGCTGGGCGT	1942
Qy	1947	GTTTACAAATAAGCCGGGTGAGCTATCCAAATGACTTCTTTGTGTAACCTGCTAGACCTCAA	2006
Db	1943	GTTCACGGAGGCTCCGAGTCACTGACCAACGACTTCTTCGTGAACCTGCTCGACATGGG	2002
Qy	2007	CACTAAATGCGAGCGACGCGATCAATCAGACAAGTTTTTTGAAGCGAGAGACTTCAAAAC	2066
Db	2003	TATCACTGGGAGCCCTTCCGACACAGATGACGGGACCTACGAGGGCAAGGATGGC---AG	2059
Qy	2067	TGGCGAAGTAAAGTTGGAGTGGCCACCGGGTAGACCTGATCTTCGGATCCAATTTCCGAGCT	2126
Db	2060	TGGCAAGGTGAAGTGACCGGACGCGCTGGACCTTGGTCTTCGGGTCCAACTCGAGTT	2119
Qy	2127	AAGAGCCCTCGCAGAAGTGTACGGCTGTGCAGATTCTGAAGAAAAGTTTTTTAAAGATT	2186
Db	2120	CGGGGCGCTTTCGAGGTCTATGGCGCGGATGACGCGCAGCCGAAGTTCGTGCGGACATT	2179
Qy	2187	TGTGAGGCCCTGGGCCAAAGTAATGACCTGGACCGGTTTGATCTG	2232
Db	2180	CGTCGCTGCTGGGCAAGGTGATGAACCTCGACAGGTTCCGAGCTG	2225

RESULT 10

US-08-852-219-1
 : Sequence 1, Application US/08852219
 : Patent No. 5922575
 : GENERAL INFORMATION:
 : APPLICANT: Cockerill, Franklin R.
 : APPLICANT: Kline, Bruce C.
 : APPLICANT: Uhl, James R.
 : TITLE OF INVENTION: Detection of Isoniazid Resistant Strains
 : TITLE OF INVENTION: of M. Tuberculosis
 : NUMBER OF SEQUENCES: 22
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Mueiting, Raasch & Gebhardt, P.A.
 : STREET: 119 No. 5922575th Fourth Street, Ste. 203
 : CITY: Minneapolis
 : STATE: MN
 : COUNTRY: USA
 : ZIP: 55401
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/852,219
 : FILING DATE: 07-May-1997
 : CLASSIFICATION: 435
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Sandberg, Victoria A.
 : REGISTRATION NUMBER: 41,287
 : REFERENCE/DOCKET NUMBER: 230.00010130
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 612-305-1226
 : TELEFAX: 612-305-1228
 : INFORMATION FOR SEQ ID NO: 1:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 2235 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: DNA
 : US-08-852-219-1

Query Match 37.2%; Score 832.4; DB 2; Length 2235;
Best Local Similarity 63.0%; Pred. No. 2.7e-247;
Matches 1389; Conservative 0; Mismatches 801; Indels 16;

QY 30 TACGTATAACACAAACACTGGCGGAAAAATGCCCTTTTACCGGAGGTTTCGCTTAAGCAAAG 89

Db	33	TACAGAAACCAACACCGGAGCGCTAGCAACGGCTGTCCCGTCGTGGGTCTATATGAAATA	92
Qy	90	TGCAGGTGGCGGCACCAAAACAGGATTTGGTGGCCCAACATGCTTCAACCTCGGCATCTT	149
Db	93	CCCCGTCGAGGCGCGGAAACAGGACTGTGTGGCCCAACCGGCTCAATCTGAAGGTACT	152
Qy	150	AGCCCAACATTATCGCTATCGGACCCAAACAGACCGGATTTTGACTATGCCGAAAGATT	209
Db	153	GCACAAAACCGGCGCTCGTGACCCGATGGTGCCTGCGCTTCGACTATGCCGCGGAGGT	212
Qy	210	TAAGAAGCTAGATCTGGCAGCGGTTAAAGAGACCTGGCAGCGCTTAATGACAGATTACAC	269
Db	213	CGCGACCATCGACGTTGACGCCCTCGACCGGSACATCGAGGAAGTGATGACCACCTCGCA	272
Qy	270	GGACTGTGGCAGCAGATTAAGGTCATTATGGCCCTTCTTTATACGCATGGCGTGCCA	329
Db	273	GCCGTGGTGGCCGCGGACTACGGCCACTACGGGCGCTGTTATCCGGATGGCGTGCCA	332
Qy	330	CAGCCCGGCACCTTACCCTATCGGTATCGGCGTGTGGCGGTGGCTCCGGCTCACACGC	389
Db	333	CGCTGCCGCGCACTACCCATCCACGACGCGCGCGCGCGCGCGCGCGCATGCAACGC	392
Qy	390	CTTGCAGCTCTCAATAGCTGGCCAGACAATGCCAATCTGATAAAGCACGCTGCTTCT	449
Db	393	GTTTCGCGCGCTTAACAGCTGGCCGACACGCCAGCTTGACAAAGGCGCGCGGCTGCT	452
Qy	450	TTGCCCATCAAAACAAATACGGTCGAAAAATCTCCTGGCGGACTCTAATGATACTCAC	509
Db	453	GTGCGCGTCAAGAAGAAGTACGCCAAGACTCTCATGGCGGACCTGATTGTTTCGCC	512
Qy	510	AGGAAACGTAGCTCTGGAACATATGGGCTTTAAACTTTTGGTTTGCAGGTGCGACAGC	569
Db	513	CGGCAACTCGCGCTGGAATCGATGGGCTTCAAGACGTTCCGGTTCCGGTTCGGCGGGT	572
Qy	570	AGATGTATGGGAGCTGAAGAAGATGATCTGSGGAGCAAAACCCGATCGCTGGGAGA	629
Db	573	CGACCAGTGGGAGCC---CGATGAGGTCTATTGGGGCAAGGAAGCCACCTCGCTCGCGCA	629
Qy	630	CAAGCGCTATGAAGGTGACCGAGAGCTCGAAAAATCCCTGGGAGCGGTACAAAATGGGACT	689
Db	630	TGACGTTACAGCGGTAAAGCGGATCTGGAGAACCCGCTGCCCGGCTGCAGATGGGCT	689
Qy	690	CATCTATGTAACCCCGAAGACCCACGGCAAGCAGACCTATCGCTGCTGCGCGTGA	749
Db	690	GATCTAGCTGAACCCGGAGGGCGGAACGGCAACCCGACCCCATGGCGCGCGGTCGA	749
Qy	750	TATTCTGAGACTTTTGGCGAATGGCAATGAATGACGAACAAACCTGGCTCTCATAGC	809
Db	750	CATTCCGGAGACGTTTCGGCCATGGCCATGAACGACGTGAAACACGGCGCTGATCGT	809
Qy	810	GGGTGGACACACTTCGGAACAAACCCATGTGTCTGGCGATGCGGAGAAATATGTGGCCG	869
Db	810	CGGCGGTACACTTTTCGTTAAGACCCATGGCG---CGGCCCGGCCCATCTGGTGGGCC	866
Qy	870	AGACCTGCGCGCCGACGATATGAAGAAATGAGCCTGGGTGGGAAAAACACCTACGCGAC	929
Db	867	CGAACCAGGCTGCTCCGCTGGACGAGATGGGCTTGGCTTGGGAAGAGCTCGTATGGCAC	926
Qy	930	CGGACAGGTGGCGATACCATCACCAGTGAAGAGGCGCTGGACCAAGACCCCTAC	989
Db	927	CGGAACCGTGAAGGACCGGATCACACCGGCATCGAGGTGCTATGGACGAACACCCCGAC	986
Qy	990	TCAATGAGCAATAACTTTTTTGAACACCTTTTGGTTACGAGTGGGAGCTTACCAAAG	1049
Db	987	GAATGGGACACAGTTTCTTCGAGATCTCTGACGGCTACGAGTGGGAGCTGACGAAGAG	1046
Qy	1050	TCCAGCTGGAGCTTATCAGTGGAAACAAAGACGGTCCGCGGGCTGGCACCATACCGGA	1109
Db	1047	CCCTGCTGGCGCTTGGCAATACACCGCAAGGAGGCGCGGTCCGCGCACCATCCCGGA	1106
Qy	1110	TGCACATGATCCCAACGAAGTCGCACGCTCCATTTATGCTCAGTACGGAACCTGGCGTGGC	1169

Db	1107	CCCCGTTGCG---GCGGGCCAGGGCGCTCCCGGACGATGCTGGCCACTGACCTCTCGCTGCGG	1167
QY	1170	CATGACCCCTGATTACGAAAAAATTTCTGACGGTACTATGAAACCCCTGATGAGTTGCG	1229
Db	1164	GGTGGATCCGATCTATGACGGGATCACCGCTGCGTGGCTGGAACACCCCGAAGTAATGGC	1223
QY	1230	AGATGCTTTGCGGAAGCATGGTACAAATCAGACACAGAGATATGGAGCCAAAGGTGCG	1289
Db	1224	CGACGAGTTGCGCAGGCGCTGGTACAGCTGATCCACGGAGCATGGGTCCCGTTGGCG	1383
QY	1290	CTACCTGGGACAGAAAGTGCCCTCAGGAAGACCTCATCTGGCAAGACCTATACCAAGATGT	1349
Db	1284	ATACCTTGGGCGCTGGTCCCAAGACAGACACCTGCTGTGGCAGGATCCGCTCCCTGCGGT	1343
QY	1350	AGCCATCCTCTTTAGACGAAAACGATATTTGAAGGCTTAAAGCCAAAATCTCTGGAATC	1409
Db	1344	CAGCCACGACCTCGTGGCGGAAGCGGAGATTTGCCAGCCTTAAAGGCCAGATCGGGGCATC	1403
QY	1410	GGGACTGACGGTAAAGCAGAGCTGGTAAAGCAGCGCATGGCTTCTGTCATCTACTTTTAGAA	1469
Db	1404	GGGATTGACTGTCTACAGCTAGTTTCGACCGCATGGCGGGCGGTCTGTTTCGTTGG	1463
QY	1470	CTCTGACAAGCGGGCGGTGCCACGGTGCAAGTATACGACTGGCCCCACAAAAGACTG	1529
Db	1464	TAGCGACAAGCGCGCGCGCAACGGTGGTGGCATCCGCTGCAGCCACAAGTCGGGTG	1523
QY	1530	GGAGTAAACAACCTCA- --GCAACTTGCAGGGTACTCAAAACACTAGAAAGTATCCA	1586
Db	1524	GGAGTCAACGACCAGCCGACGGGATCTGGCGAAGTCAITTCGACCCCTGGAGAGATCCA	1583
QY	1587	GGAGACTTTAACGAGCGCAATCAGATAAACAAGCAGTATCGTTGGCCGACCTGATTTG	1646
Db	1584	GGAGTCATTCAACTCCGCGGCGCGGGANACTCAAAAGTGTCTTTCGCCGACCTCGTGT	1643
QY	1647	GCTGGCGGCTGTGCGGGGTAGAAAAGCTGCAAAAGATGCTGGCCATGAGGTGCAGGT	1706
Db	1644	GCTCGGTGGCTGTGCCGCCATAGAGAAGCAGCAAGGGCGCTGGCCACAACATCACGTT	1703
QY	1707	GCCTTTCAACCCGGGAGCGGGATGCCCGGCTGAGCAAAACCGATGTGGAGCTTTGCA	1766
Db	1704	GCCCTTACCCCGGGCGCAGGATGGTGGCAGGAACAACACCGAGTGGATTCCTTTGCG	1763
QY	1767	AGCACTAGAGCAGCGGCTGACGGCTTTAGAAACTACATTTAAACCGGAGCATAAAGTATC	1826
Db	1764	CGTCTGGAGCCCAAGCAGCATGGCTCCGNAACTACCTCGGAAGGGCANCCGTTGCC	1823
QY	1827	CGCTGAGGAATGCTCTAGACCGGGCGCAGCTTCTGTGCTTTTCGGCACCAAGATGAC	1886
Db	1824	GGCCGAGTACATGTGCTCGACAGGGCAACCTGCTTAGCTCAGTGGCCCTCAGATGAC	1883
QY	1887	TGCTTTGTAGGCGGTATGCTGTACTGGGCACCAACTACGAGCGGTTCGACGATGGAGT	1946
Db	1884	GGTCTGGTAGTGGCTGCGGCTCTCGG- GCAAACTACAGCGCTTACCGCTGGCGGT	1942
QY	1947	GTTTTACAAATAAGCCGGGTGAGCTATCCATGACTTTTGTGTAACCTGCTAGACCTCAA	2006
Db	1943	GTTCACCGAGCCCTCCGAGTCACTGACCAACGACTCTTCGTGTGAACCTGCTCGACATGGG	2002
QY	2007	CACATAATGGGAGCCAGCGATGAATCAGACAAAGTTTTTGAAGGCAGAGACTTCAAAAC	2066
Db	2003	TATCACCTGGGAGCCCTCGCCAGCAGATGACGGGACCTACCGGGCAAGGATGGC- --AG	2059
QY	2067	TGGCGAAGTAAAGTGGAGTGGCACCGGGGTAGACCTGATCTTCGGATCCAAATCCGAGCT	2126
Db	2060	TGGCAAGGTGAAGTGGACCGGACGCGGTGGACCTGGTCTTCGGGTCCCAACTCGGAGTT	2119
QY	2127	AAGAGCCCTCGCAGAAGTGTACGGCTGTGCAGATCTGAAGAAAAGTTGTTTAAAGATTT	2186
Db	2120	GCGGGCGCTTGTGAGGTCTATGGCGCGATGACGCGCAGCGAGTTCGTGAGGACTT	2179
QY	2187	TGTGAAGCCCTGGGCCAAAGTAAATGACCTGGACCGGTTTGATCTG	2232
Db	2180	CGTGCCTGCTGGGACAAAGGTGATGAACCTCGACAGGTTTCGACGTG	2225

RESULT 11
US-08-418-782-2
; Sequence 2, Application US/08418782
; Patent No. 5658733
; GENERAL INFORMATION:
; APPLICANT: Cockerill, Franklin R.
; APPLICANT: Kline, Bruce C.
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Isoniazid Resistant Strains
; TITLE OF INVENTION: of M. Tuberculosis
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg & Woessner
; STREET: 3500 IDS Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/418,782
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D.
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 150.141US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2221 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-418-782-2

Query Match 32.3%; Score 721.8; DB 1; Length 2221;
Best Local Similarity 60.5%; Pred. No. 4.8e-213;
Matches 1335; Conservative 0; Mismatches 842; Indels 28; Gaps 8;

QY	30	TACGATATACACAACTGGCGGAAATGCCCTTTTACCGGAGGTTGCGTTAAGCAAAG	89
DB	33	TACAGAAACACACCGGAGCGCTAGCAACGGCTGTCGCGTGGTGCATATGAATA	92
QY	90	TGCAGGTGGCGGACCAAAACAGGATGGTGGCCCAACATGCTCAACCTCGGCATCTT	149
DB	93	CCCGTCGAGGCGCGGAAACAGAGCTGGTGGCCCAACCGCTCAATCTGAAGTACT	152
QY	150	ACGCCAACATTCATCGCTATCGGACCCCAACGACCGGATTTTGAATATGCCGAAGTT	209
DB	153	GCACCAAAACCGCGCGTGCCTGACCCGATGGTGGCGGCTTCGATATGCCCGGAGGT	212
QY	210	TAGAAGCTAGATCTCGGAGCGGTTAAAGAACCTGGGACGCGCTAATGACAGATTACA	269
DB	213	CGGACAGCTGCACTTTCAGCGCCCTGACGGGGACATCGAGGAAGTGATGACCACTCGCA	272
QY	270	GGACTGGTGGCCAGAGATTACGTCATTTATGCCCTTCTTTATACGATGCGGTGCA	329
DB	273	GCGGTGGCGCGGAGTACGCGCCTACGCGCGCTGTTTATCCGATGCGGTGGCA	332
QY	330	CAGCGCGGACCTACCGCTATCGGTGATGGCGGCTGGTGGCGGTGGCTCCGCGTACAGCG	389
DB	333	CGCTGGCGGACCTACCGCATCCACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	392

QY	390	CTTCGGCGCTCTCAATAGCTGGCCAGACAATGCAATCTGGATAAAGACAGCTTGTCTCT	449
DB	393	GTTCGGCGCGCTTAAACAGCTGGCCGACAAACGCAAGCTTGGCAAGCGCGCGCGCTGT	452
QY	450	TTGGCCCATCAACAAATAACGCTCGAATAATCTCTGGCGGATCTAATGATACATCAC	509
DB	453	GTGGCCGCTCAAGAGAGTACGGCAAGAGCTCTCATGGCGGAGCTGATTTTTCGC	512
QY	510	AGSAAAC--GTAGCTCTGGAACATATGGGCTTTTAAACTTTTGGTTTTCAGTGCAGACA	567
DB	513	CGSAAACCGCTGGCTCGGAATCGATGGGCTTCAAGACGCTTCGGTTTCGGCGCT	572
QY	568	GCAGATGATGGAGCCTGAAGAGATGATATCTGGGAGCAGAAACCAACGATGCTGGA	627
DB	573	CGACCACTGGGAGAC-----GATAGGCTTATTTGGGCAAGGAAGCCACCTGCTCGCG	627
QY	628	GACAAGCGCTATGAAGGTGACCGAGAGCTCGAAATCCCTGGAGCGCTACAAATGGGA	687
DB	628	GATGA---CGTTACAGCGTAAGCGATCTGGAGAACCGCTGCCCGGTGCAAGTGGGG	684
QY	688	CTCATCTATATAACCCCGAAGGACCCAAAGCGGAGCCAGACCTATCGCTGCTGGCGCT	747
DB	685	CTGATCTAGTGAACCGGAGCGCGCAACCGGACCCCATGGCGCGCGGCTC	744
QY	748	GATATTCGTGAGACTTTTGGCCCAATGGAATGACAGAAACCGTGGCTCTCATATA	807
DB	745	GACATTCGGGAGACGTTTCGGCGCATGGCCATGAAGCGCTCGAAACAGCGCGCTGATC	804
QY	808	GCGGTGGGACACACTTCGSGAAACCCATGGTGTGCGGATCGGAGAGATATGTGGC	867
DB	805	GTGCGCGTCACTTTTCGGTAAGACCCATGG---CGCCGCGCGCGCGCTGTGGTGGC	861
QY	868	CGAGACCTGCGCGCGCAGGTATTGAAGAAATGAGCTTGGGTGGAAACACCTACCGC	927
DB	862	CCGGAACCGAGGCTGCTCCGCTGGAGCAGATGGCTTGGGTGGAAGAGCTGCTATGGC	921
QY	928	ACGGGACAGGTGGGATACCATCACAGTGGAGTGAAGGCGCTGGACCAAGACCCCT	987
DB	922	ACGGAAACCGGTAAAGACCGCATCACACGCGGATCGAGTCTGATGGAGAACACCCCG	981
QY	988	ACTCAATGGAGCAATAACTTTTGTAAACCTCTTTGTACAGTGGGAGCTTACCAA	1047
DB	982	ACGAATGGGACAAACAGTTTCTCGAGATCTCTGAGCTACAGTGGGAGCTGACGAAG	1041
QY	1048	AGTCCAGCTGGAGCTTATCAGTGGAAACCAAAAGAGCGTGGCGGCTGGACCATACCG	1107
DB	1042	AGCCCTGCTGGCGCTTGGCAATACCGCCAAAGACGCGCGCTGGCGCACCATCCCG	1101
QY	1108	GATGCACATGATCCGAGCAGTGCAGCGCTCCATTTATGCTCACTAGGACCTGGCGCTG	1167
DB	1102	GACCCGTTCCG---GCGGGCGCAGGCGCTCCCGACGATGCTGCCACTGACCTCTGCTG	1158
QY	1168	CGCATGACCTGATACGAAATAATTTCTCGAGCGTACTATGAAACCCCTGATGAGTTT	1227
DB	1159	CGGCTGGATCCGATCTATGAGCGGATCACCGCTGCTGGCTGGAAACACCCCGAGGATG	1218
QY	1228	GCAGATGCTTTTCGGAAGCATGTTACAACTGACACACAGAGATATGGACCAAGAGTG	1287
DB	1219	GCCGACGAGTTCGCAAGCGCTGTACAACTGATCCACGAGCATGGTCCCGTTGCG	1278
QY	1288	CGTACTCTGGGACAGAAAGTGCCTCAGGAAGACCTCATCTGGCAAGACCTATACCAAT	1347
DB	1279	AGATACCTTGGCGCGTGGTTCGCAAGCAGACCTGCTGTGGCAGGATCCGCTCCGCTG	1338
QY	1348	GTAAGCATCTCTTGTAGACGAAACAGATATTTGAAGCCCTAAAGCCAAATCTTGAA	1407
DB	1339	GTGAGC---ACGACCTGCTCGCGGAGAGATTTGCCAGCTTAAGAGCCAGATCCGCGCA	1395
QY	1408	TCCGGGACTGACGCTAAGCGAGCTGGTAAAGCAGCGCATGGGCTTCTGCTACTCTTTAGA	1467
DB	1396	TCGGGATGACTCTCTCACAGCTAGTTTCCACCGCATGGCGCGCGCTGCTGTTCCGT	1455
QY	1468	AATCTGACAAAGCGCGCGTGCACAGCGTATACGACTGAGCGCGCGCGCGCGCGCGCG	1527

```
Db 1456 GGTAGCGAAGCGCGCGCGCCCAACGGTGGTGGCATCCCGCTGCACGCAAGTCCGG 1515
Qy 1528 TGGGAAGTAAACAACTTCAGCAACTTGGCAGGGTACTCAAAACACACTAGAGGATCCAG 1587
Db 1516 TGGGAGTCAACAGACCGCGCGGATTCGCGGAAGTCAATTCGACCCCTGAAGATCCAG 1575
Qy 1588 GAGGACTTTAACAGAGCGCAATAGATAACAAAGCATATCGTTGGCGACCTGATGTTG 1647
Db 1576 GAGTCAATCA-----CTCGCGCGGGAACATCAAAAGTGTCTTCGCGGACCTCGTGTG 1629
Qy 1648 CTGCGCGCTGTGGGGTGTAGAAAAGCTCAAAAGATGCTGGCATGAGTGCAGGTG 1707
Db 1630 CTCGGTGGCTGTGGCCACTAGAAAAGCAGCAAGCGGCTGGCCACAACTACGCGTG 1689
Qy 1708 CTTTCAACCGCGGACGAGCGGATCGCACCGCTGAGCAAAAGCATGGAAGCTTTCGA 1767
Db 1690 CCTTCAACCGCGGCGCGCGAGATGCGTGGCAGCAAAACCGAGTGGATCTTTGCC 1749
Qy 1768 GCATAGAGCAGCGGTGAGCGGCTTTAGAAACTACATTAACACCGGAGCATAAAGTATCC 1827
Db 1750 GTGCTGGAGCCCAAGGAGATGGCTTCGGAACCTACCTCGGAAAGGCAACCGTTGCCG 1809
Qy 1828 GCTGAGGAATGCTGCTAGACCGGCGGAGCTTCTGTCGGTTCGGCACCAAGAAATGACT 1887
Db 1810 CCGAGTACATCGCTGCTGCAAGGCGCAACCTGCTTACGCTCAGTGCCCTGAGATGAG 1869
Qy 1888 GCTTTGTTAGCGGTATGCTGTACTGGGCAACCACTACGACGCTTCGACGATGGAGTG 1947
Db 1870 GTGCTGTGAGTGGCTGCGGCTCCTCGGCGCAAACTACAAGCCTTACCGCTGGCGTG 1929
Qy 1948 TTTCAAAATAAGCGCGGTGAGCTATCAATGACTTCTTTGTAACCTGCTAGACCTCAAC 2007
Db 1930 TTCACCGAGGCTCCGAGTCACTGACCAACGACTTCTTCGTGAACCTGCTGACATGGT 1989
Qy 2008 ACTAAATGGCGAGCGGATGATCAGACAAAGTTTTGAAGCGCAGAGCTTCAAAACT 2067
Db 1990 ATCACTGGGAGCCTCGCCAGCAGATGACGGGACCTACCAAGGCAAGGATGGC---AGT 2046
Qy 2068 GCGAAGTAAAGTGGAGTGGCACCAGGCTAGACCTGATCTTCGGATCCAATTCGAGCTA 2127
Db 2047 GCGAAGTGAAGTGGACCGCGAGCGGCTGACCTGCTTCGGTCCAACTCGGAGTTG 2106
Qy 2128 AGACCCCTCGCAGAAAGTACGGGTGTGAGATTCCTGAAGAAAAGTTGTTAAAGATTTT 2187
Db 2107 CGGCGCTTCTGAGGCTATGCGCCGATGACGCGGAGGCAAGTTCGTGACAGATTC 2166
Qy 2188 GTGAAGGCTGGGCAAGTAATGAGCTGACCGGTTGATCTG 2232
Db 2167 GTCCTCGGTGGGACAAGTGTGTAACCTCGACAGGTTGACGTG 2211
```

RESULT 12

```
US-08-228-662-2
; Sequence 2, Application US/08228662
; Patent No. 5688639
; GENERAL INFORMATION:
; APPLICANT: COCKERILL, FRANKLIN R.
; APPLICANT: KLINE, BRUCE C.
; APPLICANT: UHL, JAMES R.
; TITLE OF INVENTION: DETECTION OF ISONIAZID RESISTANT STRAINS
; NUMBER OF INVENTION: OF M. TUBERCULOSIS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESS: SCHWEGMAN, LUNDBERG & WOESSNER, P.A.
; STREET: 3500 IDS CENTER
; CITY: MINNEAPOLIS
; STATE: MINNESOTA
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/228,662
FILING DATE: 18-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: WOESSNER, WARREN D.
REGISTRATION NUMBER: 30,440
REFERENCE/DOCKET NUMBER: 150.123US01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-339-0331
TELEFAX: 612-339-3061
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2221 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-228-662-2

Query Match 32.3%; Score 721.8; DB 1; Length 2221;
Best Local Similarity 60.5%; Pred. No. 4.8e-213;
Matches 1335; Conservative 0; Mismatches 842; Indels 28; Gaps 8;

Qy 30 TACGTATATACACAAACACTGCGGGAATGCCCTTTTACCGGAGGTTCGCTTAAGCAAG 89
Db 33 TACAGAACCAACACCGGAGCGGTAGCAAGCGGTGTCCCGTGGTGGTTCATATGAATA 92
Qy 90 TSCAGGTGGCGGCCCAAAAAAGGGATTGTGGCCCAACATGCTCAACCTCGGCATCTT 149
Db 93 CCCCGTGGAGGCGCGGAAACAGGACTGTGGCCCAACCGCTCAATCTGAAGGTACT 152
Qy 150 ACGGCACATTCATCGCTATCGGACCCAAAGACCGGATTTTACATATGCCGAAGATT 209
Db 153 GCACAAAACCGCGCGCTGCTGACCCGATGGGTGCGGCGTTCGACTATGCGCGGAGGT 212
Qy 210 TAAGAAGCTAGATCTGCGCAGCGGTTAAAAAGAGCTGGCAGCGCTAATGACAGATTACA 269
Db 213 CGCGNCCAGTCGACTTGACGCCCTTGACGCGGAGCATCGAGGAAGTGATGACCACTCGCA 272
Qy 270 GGAATGTTGGCCAGCAGATTACGGTCAATTATGGCCCTTCTTTATACGATGGCGTGCA 329
Db 273 GCCGTGTGGCCCGCCGACTACGGCCACTACGGCGCGCTGTTATCCGATGGCGTGCA 332
Qy 330 CAGCGCGGACCTACCGTATCGGTGATGCGCGTGGTGGCGGTGCTCCGCGCTCACAGCG 389
Db 333 CGCTGCGGACCTACCGCATCCACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 392
Qy 390 CTTCCGCGCTCTCAATAGCTGGCCAGACAATGCCAAATCTGATTAAGCACGCTTGTCTCT 449
Db 393 GTTCGCGCGCTTAACAGCTGGCCCGACACGCGCAGCTTGGACAAGGCGCGCGCTGCT 452
Qy 450 TTGCCCCATCAACAAAAATACGGTGCAGAAATCTCTGGCGGATCTTAATGATCTCAC 509
Db 453 GTGCGCGCTCAAGAAAGATACGGCAAGAAGCTCTCATGGCGGACCTGATGTTTTCG 512
Qy 510 AGGAAC--GTAGCTCTGGAACCTATGGCTTTTAAACTTTTGGTTTTCAGAGTGGCAGA 567
Db 513 CGGCAACCGCTGCGCTCGGAATCGATGGGCTTCAAGAGCTTCGGGTTTCGGGCTG 572
Qy 568 GCAGATGTATGGGAGCCTGAAGAAGATGTATACTGGGAGAGCAGAAACCGAATGTCTGGA 627
Db 573 CGACCACTGGGAGAC-----GATGAGGTCTATTGGGCAAGGAAGCACTGCTGCTGCG 627
Qy 628 GACAAGCGCTATGAAGGTGACCGAGAGCTCGAAAATCCCTGGAGCGGTACAAATGGGA 687
Db 628 GATGA---CGGTTACAGCGTAAGCGATCTGGAGAACCCTGCGCGCGGTGACAGATGGG 684
Qy 688 CTCATCTATGTAAACCCCGAGGACCAACGGCAAGCCAGACCCCTATCGCTGCTGCGGT 747
Db 685 CTGATCTACGTGAACCGGAGCGCGCGCAACCGCAACCGGACCCCACTGGCGCGCGGCTC 744
```

QY 748 GATATTCTGTAGACATTTTGGCGGAATGGAATGAATAGCAAGAAACCGTGGCTCTCAT 807
D 749 GATATTCTGTAGACATTTTGGCGGAATGGAATGAATAGCAAGAAACCGTGGCTCTCAT 807
D 750 GATATTCTGTAGACATTTTGGCGGAATGGAATGAATAGCAAGAAACCGTGGCTCTCAT 807
QY 808 GCGGTGTGACACACCTTCGGAACCAACCATGGTGTGCGGATGCGGAGAAATATGTGGC 867
D 809 GCGGTGTGACACACCTTCGGAACCAACCATGGTGTGCGGATGCGGAGAAATATGTGGC 867
D 810 GCGGTGTGACACACCTTCGGAACCAACCATGGTGTGCGGATGCGGAGAAATATGTGGC 867
QY 868 CGAGACCTGGCGCGCGAGGTATTGAAGAAATGAGCCTGGGTGGAACCAACCATGTACGCG 927
D 869 CGAGACCTGGCGCGCGAGGTATTGAAGAAATGAGCCTGGGTGGAACCAACCATGTACGCG 927
D 870 CGAGACCTGGCGCGCGAGGTATTGAAGAAATGAGCCTGGGTGGAACCAACCATGTACGCG 927
QY 928 ACCGACACGGTGTGCGGATGACATACCATGAGTGTAGAGGCGCTGGACCAAGACCCCT 987
D 929 ACCGACACGGTGTGCGGATGACATACCATGAGTGTAGAGGCGCTGGACCAAGACCCCT 987
D 930 ACCGACACGGTGTGCGGATGACATACCATGAGTGTAGAGGCGCTGGACCAAGACCCCT 987
QY 988 ACTCAATGAGCAATTAACCTTTTGAAGAACTCTTTGTTAGAGTGGAGCTTACCAAA 1047
D 989 ACTCAATGAGCAATTAACCTTTTGAAGAACTCTTTGTTAGAGTGGAGCTTACCAAA 1047
D 990 ACTCAATGAGCAATTAACCTTTTGAAGAACTCTTTGTTAGAGTGGAGCTTACCAAA 1047
QY 1048 AGTCAGCTGGAGCTTATCAGTGAACCAACCAAGAGCGGTGCGGGGCTGGACCAATACCG 1107
D 1049 AGTCAGCTGGAGCTTATCAGTGAACCAACCAAGAGCGGTGCGGGGCTGGACCAATACCG 1107
D 1050 AGTCAGCTGGAGCTTATCAGTGAACCAACCAAGAGCGGTGCGGGGCTGGACCAATACCG 1107
QY 1108 GATGACATGATCCAGCAAGTGCACGCTCCTATTTATGCTCACTACGACACCTGGCGTG 1167
D 1109 GATGACATGATCCAGCAAGTGCACGCTCCTATTTATGCTCACTACGACACCTGGCGTG 1167
D 1110 GATGACATGATCCAGCAAGTGCACGCTCCTATTTATGCTCACTACGACACCTGGCGTG 1167
QY 1112 GACCGGTTCG---GCGGGCGAGGCGCTCCCGACGATGCTGGCCACTGACCTCTCGCTG 1158
D 1113 GACCGGTTCG---GCGGGCGAGGCGCTCCCGACGATGCTGGCCACTGACCTCTCGCTG 1158
D 1114 GACCGGTTCG---GCGGGCGAGGCGCTCCCGACGATGCTGGCCACTGACCTCTCGCTG 1158
QY 1168 GCGATGGACCTGATTCGMAAATTTCTGACGGTACTATGAACCCCTGATGATT 1227
D 1169 GCGATGGACCTGATTCGMAAATTTCTGACGGTACTATGAACCCCTGATGATT 1227
D 1170 GCGATGGACCTGATTCGMAAATTTCTGACGGTACTATGAACCCCTGATGATT 1227
QY 1228 GCAGATGCTTTCGGAAGCAATGATACAACTGACACACAGAGATATGGACCAAGGTG 1287
D 1229 GCAGATGCTTTCGGAAGCAATGATACAACTGACACACAGAGATATGGACCAAGGTG 1287
D 1230 GCAGATGCTTTCGGAAGCAATGATACAACTGACACACAGAGATATGGACCAAGGTG 1287
QY 1288 CGCTACCTGGGACAGAGTGTCTCAGGAAGACCTATCTGTGCAAGACCTATACAGAT 1347
D 1289 CGCTACCTGGGACAGAGTGTCTCAGGAAGACCTATCTGTGCAAGACCTATACAGAT 1347
D 1290 CGCTACCTGGGACAGAGTGTCTCAGGAAGACCTATCTGTGCAAGACCTATACAGAT 1347
QY 1291 AGATACCTTGGGCGCTGTGTCGCAAGAGACCTCTGTGTGCGAGGATCGGCTCGCTGG 1338
D 1292 AGATACCTTGGGCGCTGTGTCGCAAGAGACCTCTGTGTGCGAGGATCGGCTCGCTGG 1338
D 1293 AGATACCTTGGGCGCTGTGTCGCAAGAGACCTCTGTGTGCGAGGATCGGCTCGCTGG 1338
QY 1348 GTAAGCCATCTCTTTGATAGCAAAACGATATGAAGGCTTAAAGCCCAAAATCTCGAA 1407
D 1349 GTAAGCCATCTCTTTGATAGCAAAACGATATGAAGGCTTAAAGCCCAAAATCTCGAA 1407
D 1350 GTAAGCCATCTCTTTGATAGCAAAACGATATGAAGGCTTAAAGCCCAAAATCTCGAA 1407
QY 1399 GTCAGC---ACAGCTCTGCGGAGAGAGATGTCGACCTTAAGAGCCAGATCGGGCA 1395
D 1400 GTCAGC---ACAGCTCTGCGGAGAGAGATGTCGACCTTAAGAGCCAGATCGGGCA 1395
D 1401 GTCAGC---ACAGCTCTGCGGAGAGAGATGTCGACCTTAAGAGCCAGATCGGGCA 1395
QY 1408 TCGGAGCTGACGCTAAGCGAGTGTGTAAGCAGCGATGGCTTCTGATCTACTTTTGA 1467
D 1409 TCGGAGCTGACGCTAAGCGAGTGTGTAAGCAGCGATGGCTTCTGATCTACTTTTGA 1467
D 1410 TCGGAGCTGACGCTAAGCGAGTGTGTAAGCAGCGATGGCTTCTGATCTACTTTTGA 1467
QY 1396 TCGGAGTGTGCTCTACAGCTAGTTTCGACCGCATGGCGGCGCTGCTGCTCGT 1455
D 1397 TCGGAGTGTGCTCTACAGCTAGTTTCGACCGCATGGCGGCGCTGCTGCTCGT 1455
D 1398 TCGGAGTGTGCTCTACAGCTAGTTTCGACCGCATGGCGGCGCTGCTGCTCGT 1455
QY 1458 AACTCTGACAGCGGCGGTGCCAAGGTGACGATATACGACTGGCGGCCCAAAAGAC 1527
D 1459 AACTCTGACAGCGGCGGTGCCAAGGTGACGATATACGACTGGCGGCCCAAAAGAC 1527
D 1460 AACTCTGACAGCGGCGGTGCCAAGGTGACGATATACGACTGGCGGCCCAAAAGAC 1527
QY 1456 GGTAGCGACAAGCGGCGGCGCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1515
D 1457 GGTAGCGACAAGCGGCGGCGCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1515
D 1458 GGTAGCGACAAGCGGCGGCGCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1515
QY 1528 TGGGAAGTAAACCAACCTCAGCACTTCCAGGCTACTCAAAACACTAAGAGTATCCAG 1587
D 1529 TGGGAAGTAAACCAACCTCAGCACTTCCAGGCTACTCAAAACACTAAGAGTATCCAG 1587
D 1530 TGGGAAGTAAACCAACCTCAGCACTTCCAGGCTACTCAAAACACTAAGAGTATCCAG 1587
QY 1516 TGGGAGGTCAAGACCGCGGATGCTGCGAAGGTCACTGCGACCTGGAAGAGATCCAG 1575
D 1517 TGGGAGGTCAAGACCGCGGATGCTGCGAAGGTCACTGCGACCTGGAAGAGATCCAG 1575
D 1518 TGGGAGGTCAAGACCGCGGATGCTGCGAAGGTCACTGCGACCTGGAAGAGATCCAG 1575
QY 1588 GAGGACTTTAACCGCGCAATCAGATACAAAGCAGTATCTGCGCCGACCTGATTG 1647
D 1589 GAGGACTTTAACCGCGCAATCAGATACAAAGCAGTATCTGCGCCGACCTGATTG 1647
D 1590 GAGGACTTTAACCGCGCAATCAGATACAAAGCAGTATCTGCGCCGACCTGATTG 1647
QY 1576 GAGTATCA-----CTCGGCGCGGACATCAAGTGTCTTCCGCGACCTGCTG 1629
D 1577 GAGTATCA-----CTCGGCGCGGACATCAAGTGTCTTCCGCGACCTGCTG 1629
D 1578 GAGTATCA-----CTCGGCGCGGACATCAAGTGTCTTCCGCGACCTGCTG 1629
QY 1648 CTGGCGGCTGTGCGGCTGTAGAAAAGCTGCAAGATGTGTGGCCATGAGTGTGAGT 1707
D 1649 CTGGCGGCTGTGCGGCTGTAGAAAAGCTGCAAGATGTGTGGCCATGAGTGTGAGT 1707
D 1650 CTGGCGGCTGTGCGGCTGTAGAAAAGCTGCAAGATGTGTGGCCATGAGTGTGAGT 1707
QY 1630 CTGGGTGGTGTGCGCCATAGAGAAAGCAGCAAGGCGCTGGCCACCAACATCAGGTG 1689
D 1631 CTGGGTGGTGTGCGCCATAGAGAAAGCAGCAAGGCGCTGGCCACCAACATCAGGTG 1689
D 1632 CTGGGTGGTGTGCGCCATAGAGAAAGCAGCAAGGCGCTGGCCACCAACATCAGGTG 1689
QY 1708 CTTTCAACCCGGGACGAGTGCACCGCTGAGCAACCGATGTGGAAGCTTTCGAA 1767
D 1709 CTTTCAACCCGGGACGAGTGCACCGCTGAGCAACCGATGTGGAAGCTTTCGAA 1767
D 1710 CTTTCAACCCGGGACGAGTGCACCGCTGAGCAACCGATGTGGAAGCTTTCGAA 1767
QY 1690 CCGTTCACCCCGGCGCGACGATCGTGTGCGAGCAACACGAGTGTGGAATCTTTGCC 1749
D 1691 CCGTTCACCCCGGCGCGACGATCGTGTGCGAGCAACACGAGTGTGGAATCTTTGCC 1749
D 1692 CCGTTCACCCCGGCGCGACGATCGTGTGCGAGCAACACGAGTGTGGAATCTTTGCC 1749
QY 1768 GCACCTAGACCGGCGCTGACGGCTTGAAGAACTACATTAACCGGAGCATTAAGTATCC 1827
D 1769 GCACCTAGACCGGCGCTGACGGCTTGAAGAACTACATTAACCGGAGCATTAAGTATCC 1827
D 1770 GCACCTAGACCGGCGCTGACGGCTTGAAGAACTACATTAACCGGAGCATTAAGTATCC 1827
QY 1750 GTGCTGGAGCCCAAGGAGATGGCTTCCGAACTTCTCGGAAGGGAACCGCTGGCCG 1809
D 1751 GTGCTGGAGCCCAAGGAGATGGCTTCCGAACTTCTCGGAAGGGAACCGCTGGCCG 1809
D 1752 GTGCTGGAGCCCAAGGAGATGGCTTCCGAACTTCTCGGAAGGGAACCGCTGGCCG 1809

Query Match

32.3%; Score 721.8; DB 2; Length 2221;

QY 1828 GCTGAGGAATCTCTGTAGACCGCGGCGAGCTTCTGTGCTGCTTTCGGCACCAAGATGACT 1887
D 1829 GCTGAGGAATCTCTGTAGACCGCGGCGAGCTTCTGTGCTGCTTTCGGCACCAAGATGACT 1887
D 1830 GCTGAGGAATCTCTGTAGACCGCGGCGAGCTTCTGTGCTGCTTTCGGCACCAAGATGACT 1887
QY 1810 CCGAGTATACCTGCTCTGACCAAGCGAACCTTCTTACGCTCAGTCCCTGAGATGAGC 1869
D 1811 CCGAGTATACCTGCTCTGACCAAGCGAACCTTCTTACGCTCAGTCCCTGAGATGAGC 1869
D 1812 CCGAGTATACCTGCTCTGACCAAGCGAACCTTCTTACGCTCAGTCCCTGAGATGAGC 1869
QY 1888 GCTTTTGGTAGGGGTATGCTGCTACTTGGGCAACCACTAGGAGGTTTCGAGCATGAGTG 1947
D 1889 GCTTTTGGTAGGGGTATGCTGCTACTTGGGCAACCACTAGGAGGTTTCGAGCATGAGTG 1947
D 1890 GCTTTTGGTAGGGGTATGCTGCTACTTGGGCAACCACTAGGAGGTTTCGAGCATGAGTG 1947
QY 1870 GTGCTGTGAGTGGCTGCGGCTCTCTGCGGCAAACTACAAGCGCTTACCGTGGCGGTG 1929
D 1871 GTGCTGTGAGTGGCTGCGGCTCTCTGCGGCAAACTACAAGCGCTTACCGTGGCGGTG 1929
D 1872 GTGCTGTGAGTGGCTGCGGCTCTCTGCGGCAAACTACAAGCGCTTACCGTGGCGGTG 1929
QY 1948 TTTACAAATTAACCCGGGTGAGCTATCAATGACTTCTTTTGTAAACCTGTAGACCTCAAC 2007
D 1949 TTTACAAATTAACCCGGGTGAGCTATCAATGACTTCTTTTGTAAACCTGTAGACCTCAAC 2007
D 1950 TTTACAAATTAACCCGGGTGAGCTATCAATGACTTCTTTTGTAAACCTGTAGACCTCAAC 2007
QY 1930 TTTACCCGAGGCTCTGAGTCACTGACCAACGACTTCTTCTGCTGACCTGCTGACATGGT 1989
D 1931 TTTACCCGAGGCTCTGAGTCACTGACCAACGACTTCTTCTGCTGACCTGCTGACATGGT 1989
D 1932 TTTACCCGAGGCTCTGAGTCACTGACCAACGACTTCTTCTGCTGACCTGCTGACATGGT 1989
QY 2008 ACTAAATGGGAGCAGGAGTGAATCAGACAAAGTTTGTGAAGGAGAGACTTCAAACT 2067
D 2009 ACTAAATGGGAGCAGGAGTGAATCAGACAAAGTTTGTGAAGGAGAGACTTCAAACT 2067
D 2010 ACTAAATGGGAGCAGGAGTGAATCAGACAAAGTTTGTGAAGGAGAGACTTCAAACT 2067
QY 1990 ATCACTTGGGCGCTCTGCGCAGCAGATGACGGGAGCTTACCGGCAAGATGCG---ACT 2046
D 1991 ATCACTTGGGCGCTCTGCGCAGCAGATGACGGGAGCTTACCGGCAAGATGCG---ACT 2046
D 1992 ATCACTTGGGCGCTCTGCGCAGCAGATGACGGGAGCTTACCGGCAAGATGCG---ACT 2046
QY 2068 GCGAAGTAAAGTGGAGTGGCCGCGGCTGACCTGCTGCGGATCCCAATCCGAGCTA 2127
D 2069 GCGAAGTAAAGTGGAGTGGCCGCGGCTGACCTGCTGCGGATCCCAATCCGAGCTA 2127
D 2070 GCGAAGTAAAGTGGAGTGGCCGCGGCTGACCTGCTGCGGATCCCAATCCGAGCTA 2127
QY 2047 GCGAAGTAAAGTGGAGTGGCCGCGGCTGACCTGCTGCGGATCCCAATCCGAGCTA 2106
D 2048 GCGAAGTAAAGTGGAGTGGCCGCGGCTGACCTGCTGCGGATCCCAATCCGAGCTA 2106
D 2049 GCGAAGTAAAGTGGAGTGGCCGCGGCTGACCTGCTGCGGATCCCAATCCGAGCTA 2106
QY 2128 AGAGCCCTGCGCAAGTGTACGGCTGTGCGAGATTTCTGAAGAAAAGTTTCTTAAAGATTTT 2187
D 2129 AGAGCCCTGCGCAAGTGTACGGCTGTGCGAGATTTCTGAAGAAAAGTTTCTTAAAGATTTT 2187
D 2130 AGAGCCCTGCGCAAGTGTACGGCTGTGCGAGATTTCTGAAGAAAAGTTTCTTAAAGATTTT 2187
QY 2107 CCGGCGCTGTGCGAGTCTATGCGCGGATGACGCGCAGCGAGTTCGTGACAGATTTC 2166
D 2108 CCGGCGCTGTGCGAGTCTATGCGCGGATGACGCGCAGCGAGTTCGTGACAGATTTC 2166
D 2109 CCGGCGCTGTGCGAGTCTATGCGCGGATGACGCGCAGCGAGTTCGTGACAGATTTC 2166
QY 2188 GTGAAGGCTTGGGCAAGTAAATGAGCTGACCTGCGGAGCTGCTGAGTCTG 2232
D 2189 GTGAAGGCTTGGGCAAGTAAATGAGCTGACCTGCGGAGCTGCTGAGTCTG 2232
D 2190 GTGAAGGCTTGGGCAAGTAAATGAGCTGACCTGCGGAGCTGCTGAGTCTG 2232
QY 2167 GTGCTGTGCGTGGGCAAGTGAACCTGCGAGCTGCGAGGTTGCGAGCTG 2211
D 2168 GTGCTGTGCGTGGGCAAGTGAACCTGCGAGCTGCGAGGTTGCGAGCTG 2211
D 2169 GTGCTGTGCGTGGGCAAGTGAACCTGCGAGCTGCGAGGTTGCGAGCTG 2211

RESULT 13

US-08-852-219-2
; Sequence 2, Application US/08852219
; Patent No. 5922575
; GENERAL INFORMATION:
; APPLICANT: Cockerill, Franklin R.
; APPLICANT: Kline, Bruce C.
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Isoniazid Resistant Strains
; TITLE OF INVENTION: of M. Tuberculosis
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Muehling, Raasch & Gebhardt, P.A.
; STREET: 119 No. 5922575th Fourth Street, Ste. 203
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,219
; FILING DATE: 07-May-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sandberg, Victoria A.
; REGISTRATION NUMBER: 41,287
; REFERENCE/DOCKET NUMBER: 230.00010130
; TELEPHONE: 612-305-1226
; TELEFAX: 612-305-1228
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2221 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-852-219-2

Best Local Similarity 60.5%; Pred. No. 4.8e-213;
Matches 1335; Conservative 0; Mismatches 842; Indels 28; Gaps 8;

QY	30	TACGTATTAACACAACTGGCGGAAATGCCCTTTTACCAGAGTTCCGCTTAAACAAAG	89
Db			
Db	33	TACAGAAACACACCGGAGCGGTAGCAAGGGCTGCCGTGCTGGGTATATGAAATA	92
QY	90	TGCAGGTGGCGSACCAAAACAGGATGCTGGCCCAACATGCTCAACCTCGGCATCTT	149
Db			
Db	93	CCCGTGCAGGGCGGGAACAGGACTGGTGGCCCAACCGGCTCAATCTGAAGGTACT	152
QY	150	ACGCCAATTCATCGCTATCGGACCCAAAGACCCGGATTTTGACTATGCGGAAGACTT	209
Db			
Db	153	GCACAAACCCGCGCTGCTGACCCGATGGGTGGCGCTTCGACTATGCGCGGAGGT	212
QY	210	TAAAGAGTAGATGCGAGCGGTTTAAAGAGCCTGCGACGCTAATGACAGATTACAA	269
Db			
Db	213	CGCAGCAGTAGCTTACGCGCCTGACGCGGAGATCGAGGAAGTAGACCACTCGCA	272
QY	270	GGACTGGTGGCCAGCAGATTACGCTCATATTATGSCCCCTTCTTTATACGCATGGGTGCA	329
Db			
Db	273	GCCGTGGTGGCCCGGCACTACGGCCACTACGGGCGGCTGTTTATCCGGATGGCGTGGCA	332
QY	330	CAGCGCGGCACTACCGTATCGGTGATGGCGGTGGTGGCGTGGCTCCCGCTCACAGCG	389
Db			
Db	333	CGTCCGCGCACTACCGCATCCACAGCGCGCGGCGCGCGGGGGCGCATGCGAGCG	392
QY	390	CTTCGCGCTCTCAATAGCTGGCCAGACAAATGCCAATCTGGATAAAGCAAGCTTCTCT	449
Db			
Db	393	GTTCCGCGCGGTTAACAGCTGGCCGACACAGCGCAGCTTGGCAAGCGCGCGGCTGCT	452
QY	450	TTGCGCCATCAACAAATAACGCTCGAAATCTCTCTGGCGGATCTAATGATACAC	509
Db			
Db	453	GTGCGCGCTCAAGAAGAAGTACGCCAAGAGCTCTCATGGGCGGACCTGATGTTTTCG	512
QY	510	AGGAAC--GTAGCTCTGGAACATATGGGCTTTTAAACTTTTGGTTTTCAGGTGGCAGA	567
Db			
Db	513	CGGCAACCGCTCGCCTCGGAATCGATGGCTTCAAGAGCTTCGGGTTTCGGCTCGGCGT	572
QY	568	GCAGATGATGGAGCTTGAAGAAGATGTATACGTGGGAGCAGAAACCGAATGGCTGGGA	627
Db			
Db	573	CGACCACTGGGAGACC-----GATGAGGTCTATTGGGCAAGGAAGCACTTGCTCGGC	627
QY	628	GACAAGGCTTATGAAGTGCAGAGAGCTCGAAATCCCTGGGAGCCGTACAAATGGGA	687
Db			
Db	628	GATGA---CGGTTACACGTAAGGATCTGGAGAACCCGCTGGCCGGGTGCAGATGGG	684
QY	688	CTCATCTATGTAACCCCGAAGCACCAAGCCAGACCCATCGCTGCTGCTGCGCT	747
Db			
Db	685	CTGATCTACGTGAACCGGAGGCGCGAAGCGCAACCCGACCCCATGGCGCGCGGCTC	744
QY	748	GATATTCGTGAGACTTTTGGCCGAATGGCAATGAATGACGAAGAACCGGTGCTCTCATA	807
Db			
Db	745	GACATTCGGAGAGCTTTTCGGCATGGCCATGAACGACCTGAAACAGCGCGGCTGATC	804
QY	808	CGGGTGGACACACTTCGAAACCCATGGTGTGCTGCCATCGGAGAAATATGGGGC	867
Db			
Db	805	GTGCGCGGTACACTTTTCGTTAGACCCATGG---CGCCGCGCGCGCGCATCTGGTCGG	861
QY	868	CGAGAGCCTCCCGCGCAGGTATTGAAGAAATGAGCTGGGGTGGAAACACACTAGGC	927
Db			
Db	862	CCCGAACCCGAGGCTGCTCGCTGGAGCAGATGGGCTTGGGCTTGGAGAGCTCGATGGC	921
QY	928	ACCGGACAGGTCGGATACCATCACAGTGGACTAGAAGGCGCCTGGACCAAGACCCCT	987
Db			
Db	922	ACCGGAACCGGTAAAGACGCGATCACAGCGGCATCGAGTCTGATGGAGCAACCCCG	981
QY	988	ACTCAATGGAGCACTTATTTTGAACCTCTTTGGTTACGAGTGGGAGCTTACAAA	1047
Db			
Db	982	ACGAATGGGACAACTTTTCTCGAGATCTCTACGGCTACGAGTGGGAGCTGAGGAAG	1041
QY	1048	AGTCCAGCTGGAGCTTATCAGTGAACCAACAAAGACGGTGGCGGGCTGSCACCATACCG	1107
Db			

Db	1042	AGCCCTGCTGGCGCTTGGCAATATACACCGCCAGGACGGCGGCTGCCGACCACTATCCG	1101
QY	1108	GATGCACATGATCCCGACCAAGTCGCACACGCTCATTTATGCTCCTCCTACACGACCTGCGCGTG	1167
Db			
Db	1102	GACCGTTTCG---CGGGCCAGGGGCTCCCGACGATGCTGGCCACTGACCTCTCGGCTG	1158
QY	1168	CGATGGACCTGATTTAGCAAAAATTTCTCGAGGGTACTATGAAAAACCTGATGAGTTT	1227
Db			
Db	1159	CGGTGGATCGGATCTATGAGCGGATCATCGGCTGCTGGCTGGAACACCCCGAGGAATTG	1218
QY	1228	GCAGATGCTTTTCGCAAGAGCATGTTACAACTGACACACAGATATGGGACCAAGGTG	1287
Db			
Db	1219	GCCGACGAGTTTCGCAAGGCTTGGTACAACTGCTCCACCGAGATATGGTCCGCTTTCG	1278
QY	1288	CGCTACCTGGGACCAAGAGTGCCTCAGGAACACCTCATCTGGCAAGACCTTATACACAT	1347
Db			
Db	1279	AGATACCTTGGGCGCTGGTCCCCNAGCAGACCTGCTGTGGCAGGATCCGGTCCCTCGG	1338
QY	1348	GTAAGCCATCCTCTTGTAGACGAAACGATATTGAAGGCCCTPAAAGCCCAAACTCCTGGAA	1407
Db			
Db	1339	GTCAGC---ACGACCTCTGTCGCGGAAGCAGATTGCCAGCCTTAAGAGCCAGATCCGGSCA	1395
QY	1408	TCGGGACTGACGCTAAGGAGCTGTTAAGCAGCGCATGGGCTTCTGCACTACTTTTAGA	1467
Db			
Db	1396	TCGGGATGACTGCTCAGAGCTAGTTTCGACCGCATGTCGCGAAGGCTCATTCGCAACCTG	1455
QY	1468	AACCTCTGACAAAGCGCGGCTGCCAAGGTGCAGCTATACACTGGCCCAACAAAGAC	1527
Db			
Db	1456	GGTAGCAGCAGCGCGGCGCCACCGTGGTGGCATCCGCTTGCAGCCACAAAGTCGGG	1515
QY	1528	TGGGAAGTAAACAACCTCAGCAACTTGGCCAGGCTACTCAAAACACTAGAAAGTCTCCAG	1587
Db			
Db	1516	TGGGAGTCAACGACCCCGCGGATTCGCGAAGGTCATTCGCAACCTGGAAGAGATCCAG	1575
QY	1588	GAGCACTTTAACACAGGCGCAATCAGATAACAAACAGCATGCTTGGCCGACCTGATGTG	1647
Db			
Db	1576	GAGTCATTC-----CTCGCGCGGGAACATCAAAAGTGTCTTCGCGGACCTCGTCTG	1639
QY	1648	CTGCGCGCTGTGCGGCTGTAGAAAAGCTGCAAAAGATGCTGCGCTTTCGGCACCAAGTGTG	1707
Db			
Db	1630	CTCGTGGCTGTGCGCCACTAGAGAAAGCAGCAAGGGGCTGGCCACACATCACCGTG	1689
QY	1708	CTTTTAAACCGGACGAGCGGATGCGCACCGCTGAGCAAAACCGATGTGGAGCTTTCGAA	1767
Db			
Db	1690	CCCTTTCACCCCGCGCGCGCATGCGTTCGCGAGCAACAAACCGACGCTGGAATCCTTTGCC	1749
QY	1768	GCATGACGACGCGGCTGACGCTTTAGAAACTACATTTAAACCGGAGCATAAAGTATCC	1827
Db			
Db	1750	GTGCTGAGGCCCAAGGCGAGTGGCTTCGGAACCTTACCTCGGAAAGGGCAACCGTTGCCGG	1809
QY	1828	GCTGAGAAATGCTCTGACCGCGCGCAGCTTCTGCTGCTTTCGGCACCAAGAACTGACT	1887
Db			
Db	1810	CCGAGTACATCGCTGCTCGACAAAGGCAACCTGCTTACGCTCAGTGGCCCTGAGATGAGC	1869
QY	1888	GCTTTGTGAGGCGTATGCTGTACTGGGCAACCACTTACGACGTTTTCGAGCATGGAGTG	1947
Db			
Db	1870	GTGCTGTAGTGGCTTGGCGCTTCCGCGCAAACTTACAAAGGCTTACCGCTGGGGGTG	1929
QY	1948	TTTACAAATTAAGCGGGTACGCTATCCAATGACTCTTTTGTAAACCTGTGACCTCAAC	2007
Db			
Db	1930	TTACCGAGGCTCCCGAGTCACTGACCAACGACTTCTTCGTGAACTGTCTGACATGGGT	1989
QY	2008	ACTTAAATGGGCGGACGAGTGAATCAGCAAAAGTTTTCGAGCGAGAGACTTCAAACT	2067
Db			
Db	1990	ATCACCTGGAGCCCTCGCCAGCAGATGACGGACCTTACCAGGCGAAGGATGGC---AGT	2046
QY	2068	GGCGAAGTAAAGTGGAGTGGCACCCGGGTAGACCTGATCTTCGGATCCCAATCCGAGCTA	2127
Db			
Db	2047	GGCAAGTGAAGTGGACCGGCGCGCTGGACCTGCTCTTCGGGTCCCACTCGAGTTG	2106
QY	2128	AGAGCCCTCCCAAGTGTACGCTGTGCAGATTTCTGAAGAAAAGTTTGTAAAGATTTT	2187
Db			
Db	2107	CGGGCGCTTGTGAGGCTTATGCGCCGATGACGCGGCGAGGCGAAGTTCGTGACAGGATTC	2166


```

QY 1348 GTAAGCCATCTCTTGTAGAGCAAAACGATATTCAGGCGCTTAAAGCCAAATCTTGAA 1407
Db 3308 GTCAGC---ACGACCTCGTGGCGGACGAGATTCAGGCGCTTAAAGCCAGATCCGGCA 3364
QY 1408 TCGGAGCTACGGTAAGCAGCTGGTAAAGCAGCGCATGGGCTCTGTCATCTACTTTTGA 1467
Db 3365 TCGGATGACTGCTCAGACTAGTTTGCAGCCGATGGGGGGGGGCTGCTGCTCCGT 3424
QY 1468 AACTCTGCAAGCGCGGCTGCGACGGTGCAGTATAGCACTTGGCCCGCCACAAAAGAC 1527
Db 3425 GGTAGCGACAAGCGCGCGGCGCAACGGTGTGTCATCGGCTCGAGCCACACTCGG 3484
QY 1528 TGGGAAGTAACAACCTCAGCACTTGCAGGCTACTCAAAACACATAGAGGTATCCAG 1587
Db 3485 TGGGAGGTCAACGACCGCGGCGGATCTGCGCAAGGTCTATTCGCAACCTGAGAGATCCAG 3544
QY 1588 GAGGACTTTAACAGGCGCAATCAGATAACAAGCAGTATCGTTGGCGGACCTGATGTG 1647
Db 3545 GAGTCATCA-----CTCGCGCGGGAACATCAAAAGTGTCTTCGCGGACCTCGTCTG 3598
QY 1648 CTGGCGGCTGTGGGGGTGTAGAAAAGCTGCAAAAGATGCTGGCCATGAGGTGAGGTG 1707
Db 3599 CTGGTGTGCTGTGGCGGCTAGAGAAAGCAGCAAGCGCGCTGGCCCAACATCAGCGTG 3658
QY 1708 CCTTCAACCGGCGGACGAGCGGATCCACCGCTGAGCAAAACCGATGGAAGCTTTCGAA 1767
Db 3659 CCCTTTACCGCGGCGGCGGCGAGTGTGCTGCGCAAGCAAAACCGAGCTGGAATCTTTGCC 3718
QY 1768 GCACTAGAGCAGCGGCTGACGGCTTTAGAACTTACATTAACCGGAGCATAAAGTATCC 1827
Db 3719 GTGCTGAGCGGCGGAGTGTGCTTCCGAACACTACCTCGGAAGGCGCAACCTTGGCGG 3778
QY 1828 GCTGAGGAATGCTGTAGACCGGCGGAGCTTCTGTGCTTTGGGCGGAGGAAAGTACT 1887
Db 3779 CCGAGTACATCGTGTGCTGCAAGGCGGAACCTGCTTACGCTCAGTGCCCTGAGATGACG 3838
QY 1888 GCTTTGTAGCGGTATGCTGTACTGGGCGGCAACACTACGACGTTCCGAGCATGGAGTG 1947
Db 3839 GTGCTGTAGTGGGCTTGGCGGCTGCTGCGGCGCAAACTACAAGCGCTTACCGCTGGGCTG 3898
QY 1948 TTTACAAATGAAGCGGCTGAGCTATCAATGACTTCTTTTAAACCTGCTAGACCTCAAC 2007
Db 3899 TTAACGAGGCTCCGAGTCACTGACCAACGACTTCTTCTGGAACCTGCTGACATGGGT 3958
QY 2008 ACTAATGGCGGCGGAGGATGATGACAAAGTCTTTTGAAGCGAGAGACTTCAAAACT 2067
Db 3959 ATCACTGGGAGCGCTCGGAGCAGATGACGGGAGCTTACGAGGCGCAAGGATGCG---AGT 4015
QY 2068 GCGGAGTAAAGTGGAGTGGCAGCGGCTAGACTGATCTTCGATCCCAATTCGAGCTA 2127
Db 4016 GGCAGGTGAGTGGACCGGCGGCTGGAGCTGTGCTTCGGGTCCAACTCGGAGTTG 4075
QY 2128 AGAGCCCTCGCAGAGTGTACGGCTGTGCAAGATTTCTGAAGAAAAGTTTGTAAAGATTTT 2187
Db 4076 CGGCGCTTCTCGAGGTCTATGCGCGGATGACGCGGAGCGGAGTTCGTGACAGGATTC 4135
QY 2188 GTGAAGGCTTGGCGGCAAAAGTAATGGAGCTTGGACCGGTTTGTATGTC 2232
Db 4136 GTCGCTGCGTGGGACAGAGTGTGATGAACCTCGACAGGTTTCGACGTG 4180

```

RESULT 15

```

US-09-082-614A-45
; Sequence 45, Application US/09082614A
; Patent No. 6124098
; GENERAL INFORMATION:
; APPLICANT: Heym, Beate
; APPLICANT: Cole, Stewart
; APPLICANT: Young, Douglas
; APPLICANT: Zhang, Ying
; APPLICANT: Honore, Nadine
; APPLICANT: Telenti, Anello
; APPLICANT: Bodmer, Thomas

```

```

; TITLE OF INVENTION: Rapid Detection of Antibiotic Resistance
; TITLE OF INVENTION: in Mycobacterium Tuberculosis
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/09/082.614A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/313,185
; FILING DATE: 12-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 02356.0068-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4795 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-082-614A-45

```

```

Query Match 32.3% Score 721.8; DB 3; Length 4795;
Best Local Similarity 60.5%; Pred. No. 7.2e-213;
Matches 1335; Conservative 0; Mismatches 842; Indels 28; Gaps 8;

QY 30 TAGGTATACACAAACACTGGCGGAAATGCCCTTTTACCGGAGGTTTCGCTTAAGCAAG 89
Db 2002 TACAGAAACACACCGGAGCGCTAGCAACGGCTGTCGCGTCTGGTTCATATGAATA 2061
QY 90 TGCAGGTGGCGGCGGCAACAAACAGGGATTTGGTGGCCCAACATGCTCAACCTCGGCATCTT 149
Db 2062 CCGCGTGGAGGGGGGCGGAAACAGGACTGGTGGCCCAACCGGCTCAATCTGAAGGTACT 2121
QY 150 ACGCAACATTCATCGCTATCGGACCCCAACAGACCGGATTTTGACTATCCGAAGAGTT 209
Db 2122 GCACCAAAACCGGCGGCTCGCTGACCGATGGGTGGCGGCTTGCATATGCCGCGAGGT 2181
QY 210 TAGAAGCTAGATCTGGCAGCGGTTAAAGAGGACCTGGGCGGCTAATGACAGATTCACA 269
Db 2182 CGGACCGAGTCGACTTGACGCCCTGACGCGGAGATCGAGGAAAGTGATGACCACTCGCA 2241
QY 270 GGACTGGTGGCGGAGCAGATTAACGGTCAATTTATGGCCCTTCTTTATAGCGATGGGTGCA 329
Db 2242 GCGGTGGTGGCGGCGGAGTACGGCCACTACGGGCGGCTGTTTATCGGATGGCGTGCA 2301
QY 330 CAGCGCGGCGGCGGAGTACCGTATCGGTGATGGCGTGGTGGGCTGGCTCCGGCTCACAGCG 389
Db 2302 CGCTGGCGGCGGAGTACCGCATCCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2361
QY 390 CTTTGGCGGCTCTCAATAGCTGGCGGAGCAATGCCAATCTGGATAAAGACAGCTTGTCTT 449
Db 2362 GTTTCGCGGCTTAAACAGCTGGCGGAGCAACGCGAGCTTGGACAAAGCGCGCGGCTGCT 2421
QY 450 TTGGCCCATCAACAAACAAATACGGTGGAAATCTCTCTGGGCGGATCTAATGATACCTAC 509

```


Db 2422 GTGGCCGTCAGAAAGTACGGCAAGAAGCTCTCATGGCGGACCTGATTGTTTCGC 2481
QY 510 AGGAAAC--GTAGCTCTGGAACATATGGGCTTTAAAACCTTTTGGTTTTCGAGGTGGCAGA 567
Db 2482 CGGCAACCGCTGCGCTCGGAATCGATGGCTTCAAGAGTTCGGGTTCGGCTTCGGGGGT 2541
QY 568 GCAGATCTATGGGACCTGGAAGAAGATCTATACCTGGGAGCAGAAACCGAATGSGCTGGGA 627
Db 2542 CGACCAGTGGGAGACC-----GATGAGGTCTATTGGGGCAAGGAAGCCACCTGTGCTCGGC 2596
QY 628 GACAAGCGCTATGAAGGTGACCGGAGCTCGAAATCCCTGGGAGCGGTACAAATGGGA 687
Db 2597 GATGA---CGGTTACAGCGTAAGCGATCTGGAGAACCGCTGGCGCGGTGCAGATGGGG 2653
QY 688 CTCATCTATGTAACCCCGAAGGACCCAAACGGCAGACCCCTATCTGCTGCTGCGGT 747
Db 2654 CTGATCTACGTGAACCCCGGAGGCGCGAAACGCAACCCGACCCCATGGCCGCGCGGT 2713
QY 748 GATATCTGTGAGACTTTTGGCGGAATGCAATGAATGACGAAGAACCGTGGCTCTCAT 807
Db 2714 GACATTCGCGAGACGTTTCGGCGCATGGCCATGAACGACGTCGAAACAGCGCGGTGATC 2773
QY 808 CGGGGTGACACACTTCGGAATAACCCATGCTGCGGATGGGAGAAATATGTGGC 867
Db 2774 GTCGCGGTACACTTCCTGAGTTCGGAAGCCATGG---CGCGCGCGCGCGGTGCTGCGC 2830
QY 868 CGAGAGCTCGCGCGCAGGTATTGAAGAAATGAGCTGGGTGGAAACACCTACGCG 927
Db 2831 CCCGAACCGAGGCTGCTCCGCTGGAGCAGATGGGCTTGGGCTGGAAGAGCTCGTATGGC 2890
QY 928 ACCGACACGTCGGGATACATCACCAGTGGACTAGAGGGCGCTGACCAAGACCCCT 987
Db 2891 ACCGGAACGGTAAGGACGCGATCACCAGCGGCATCGAGGTGCTATGGACGAACACCCCG 2950
QY 988 ACTCAATGGAGCAATACTTTTGAACACCTCTTTGGTTACGAGTGGGAGCTTACCAA 1047
Db 2951 ACGAAATGGGACAACAGTTTCTCGAGATCTCTGACGCTAGAGTGGGAGCTGACGAAG 3010
QY 1048 AGTCAGCTGAGCTTATCAGTGGMAACCAAGACGCTGCGGGGTGGCAACCATACG 1107
Db 3011 AGCCCTGTGCGCTTGCAATACACCGCAAGGACGGCGCGGTGCGCGACCATCCG 3070
QY 1108 GATGCACATGATCCAGCAAGTCCGACGCTCCATTTATGCTCACTACGAGCTGGCGGT 1167
Db 3071 GACCGGTTG---CGGGCCAGGGCGTCCCGAGAGTCTGGCCACTGACCTCTCGCTG 3127
QY 1168 CGCATGGACCTGATACGAAAAATTTCTCGAGGTACTATGAAACCCCTGATGAGTTT 1227
Db 3128 CGGGTGGATCCGATCTATGAGCGGATCAGCGCTCGCTGGCTGGAACACCCCGAGGAATTG 3187
QY 1228 GCAGATGCTTCGCGAAGCATGGTACAACTGACACACAGAGATATGGGACCAAGGTG 1287
Db 3188 GCCGACGAGTTCGCAAGGCGTGGTACAACTGATCCACCGAGACATGGGTCCCGTTCG 3247
QY 1288 CGCTACCTGGGACAGAGTGCCTCAGGAAGACCTCTCTGCGAAGACCTTATACCAT 1347
Db 3248 AGATACCTTGGGCGCTGGTCCCAAGCAGACCTTCTGTGGCAGGATCCGGTCCCTCGC 3307
QY 1348 GTAAGCCATCTCTTGTAGAGGAAACGATATTGAAGGCCCTAAAGCCCAAAATCCTGGAA 1407
Db 3308 GTCAGC---ACGACCTCGTCGCGAAGCAGATTGCCACCTTTAAGAGCCAGATCGGGCA 3364
QY 1408 TCGGAGCTGAGCGGTAAAGCGAGCTGGTAAAGCAGGCGATGGGTCTTGCACTACTCTTTAGA 1467
Db 3365 TCGGATTTGACTGTCTACAGCTAGTTTCGAGCCGATGGGCGGCGGTCTGCTGCTCGT 3424
QY 1468 AACTCTGACAGCGCGCGGTGCCAACGGTGCAGTATACACTTGGCCCGCCCAAAAAC 1527
Db 3425 GGTAGCGAAGCGCGCGCGGCGCAACGGTGGTGCATCCGCTCGACGCCAACCTCGGG 3484
QY 1528 TGGGAAGTAAACCCCTCAGCAACTTGCAGGGGTACTCAAAACACTAGAAAGGTATCCAG 1587
Db 3485 TGGGAGGTCAACGACCCCGACGGATCTCGGCAAGGTCAATTCGCACCCCTGAAGAGATCCAG 3544

QY 1588 GAGGACTTTTAACGAGCGCCAAATCAGATAACAAGCAGTATCGTTGGCCGACCTGATTGTG 1647
Db 3545 GAGTCATTCA-----CTCGGCGCGGAACATCAAAAGTGTCTTTCGCCGACCTCTGCTG 3598
QY 1648 CTGGCCGCTGTGCGGGTGTAGAAAAGCTTGAANAAGATGCTGGCCATGAGGTGCGAGGTG 1707
Db 3599 CTCGGTGGCTGTGGCCCACTAGAGAAAGCAGCAAGGCGGCTGGCCCAACAATCACGCTG 3658
QY 1708 CCTTTCAACCCGGGACGAGGATGCCCGCTGAGCAACCGATGTGGAAGCTTTTCGAA 1767
Db 3659 CCCTTACCCCGGCGCCCGACGATGCGTTCGAGGAACAAACGAGCTGGAATCCTTTGCC 3718
QY 1768 GCATCTAGACCGAGCGGTGACGGCTTTAGAAACTACATTAACCCGAGCATAAAGTATCC 1827
Db 3719 GTGCTGGAGCCCAAGGCAAGATGGCTTCCGMAACTACCTCGGAAAGGCAACCGTTGCCGG 3778
QY 1828 GCTGAGGAATGCTCGTAGACCGGCGCAGCTTCTGTCGCTTTCGGCACCAAAATGACT 1887
Db 3779 CCGAGTACATCGCTCTCGAAGGCGAACCTGTACGCTCAGTGCCCTGAGATGAG 3838
QY 1888 GCTTTGTAGGCGGTATGCTGTACTGGGCACCAACTACGACGTTTCGACGATGGAGTG 1947
Db 3839 GTGCTGGTAGTGGCTCGGCTCTCGCGGCAAACTACAGCCGTTACCGCTGGCGGTG 3898
QY 1948 TTTACAAATAAGCCGGGTACGCTATCCAAATGACTTCTTTGTAAACCTGCTAGACCTCAAC 2007
Db 3899 TTCACCGAGGCTCCGAGTCACTGACCAACGACTTCTTCGTGAACCTGCTCGACATGGT 3958
QY 2008 ACTAATGGCGAGCAGCGATGAATCAGACAAGTTTTTTGAAGGCGAGACTTCAAACT 2067
Db 3959 ATCACTGGGAGCCCTCGCCAGCAGATGACGGGACCTACCGGGCAAGGATGGC---AGT 4015
QY 2068 GCGGAAGTAAAGTGGAGTGGCACCGGGTAGACCTGATCTTCGGATCCAATTCGAGCTA 2127
Db 4016 GGCRAAGTGAAGTGGACCGGACCGCTGGACCTGCTTTCGGGTCCAACCTCGAGTTG 4075
QY 2128 AGAGCCCTCGCAGAAAGTGTACGGCTGTGCAGATTCTGAAGAAAAGTTTGTAAAGATT 2187
Db 4076 CGGGCGCTTGTGAGGTCTATGCGCCGATGACGCGCAGGCGAAGTTCTGTGACAGGATTC 4135
QY 2188 GTGAAGGCTCGGCCAAAGTAAATGACCTGGACCGGTTTGATCTG 2232
Db 4136 GTCGCTCGTGGCAAGGTGATGAACCTCGACAGGTTCGAGGTG 4180

Search completed: October 8, 2003, 15:34:05
Job time : 137.323 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2003, 12:53:54 ; Search time 572.431 Seconds
(without alignments)
10144.788 Million cell updates/sec

Title: US-09-884-889-7
Perfect score: 2238
Sequence: 1 atggaataatcaaacactc.....accggtttgatctgaataa 2238

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0.

Searched: 1731049 seqs, 1297405648 residues

Total number of hits satisfying chosen parameters: 3462098

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications, NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2238	100.0	2238	10	US-09-884-889-7
2	844.4	37.7	82993	15	US-10-080-170-645
3	672.2	30.0	2262	10	US-09-884-889-5
4	251.8	11.3	620	11	US-09-940-925A-145
5	251.8	11.3	620	11	US-09-940-925A-149
6	251.8	11.3	620	11	US-09-941-193A-145
7	251.8	11.3	620	11	US-09-941-193A-149
8	250.2	11.2	620	11	US-09-940-925A-143
9	250.2	11.2	620	11	US-09-940-925A-146
10	250.2	11.2	620	11	US-09-940-925A-147
11	250.2	11.2	620	11	US-09-940-925A-150
12	250.2	11.2	620	11	US-09-941-193A-143
13	250.2	11.2	620	11	US-09-941-193A-146
14	250.2	11.2	620	11	US-09-941-193A-147
15	250.2	11.2	620	11	US-09-941-193A-150
16	248.6	11.1	620	11	US-09-940-925A-144

c 17	248.6	11.1	620	11	US-09-940-925A-148
18	248.6	11.1	620	11	US-09-941-193A-144
c 19	248.6	11.1	620	11	US-09-941-193A-148
20	157.2	7.0	391	10	US-09-825-574-1
21	157.2	7.0	391	12	US-09-882-945A-1
22	155.6	7.0	391	10	US-09-825-574-3
23	155.6	7.0	391	10	US-09-825-574-4
24	155.6	7.0	391	12	US-09-882-945A-3
25	155.6	7.0	391	12	US-09-882-945A-4
26	154	6.9	391	12	US-09-825-574-2
27	154	6.9	391	12	US-09-882-945A-2
28	38.8	1.7	4301	10	US-09-917-800A-1666
29	38.8	1.7	4301	12	US-10-205-194-126
c 30	37.4	1.6	9025608	14	US-10-156-761-1
31	36.8	1.6	1049	12	US-10-140-472-358
32	36.8	1.6	1049	12	US-10-141-761-358
33	36.8	1.6	1049	12	US-10-142-885-358
34	36.8	1.6	1049	12	US-10-158-790-358
35	36.8	1.6	1049	14	US-10-123-155-358
36	36.8	1.6	1049	15	US-10-146-731-358
c 37	36.8	1.6	10470	8	US-08-781-986A-20
38	36	1.6	520	14	US-10-184-644-332
39	36	1.6	520	14	US-10-184-634-332
c 40	36	1.6	3523	14	US-10-108-877-1
41	35.8	1.6	548	13	US-10-027-632-186215
42	35.8	1.6	888	12	US-10-140-472-544
43	35.8	1.6	888	12	US-10-141-761-544
44	35.8	1.6	888	12	US-10-142-885-544
45	35.8	1.6	888	12	US-10-158-790-544

ALIGNMENTS

RESULT 1

US-09-884-889-7
; Sequence 7, Application US/09884889
; Patent No. US20020102680A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: ROBERTSON, Dan
; APPLICANT: SANVAL, Indrajit
; APPLICANT: ADHIKARI, Robert
; TITLE OF INVENTION: CATALASES
; FILE REFERENCE: DIVER1100-4
; CURRENT APPLICATION NUMBER: US/09/884,889
; CURRENT FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 09/412,347
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: US 08/951,844
; PRIOR FILING DATE: 1997-10-16
; PRIOR APPLICATION NUMBER: US 08/674,887
; PRIOR FILING DATE: 1996-07-03
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 2238
; TYPE: DNA
; ORGANISM: Microscilla furvescens
US-09-884-889-7

Query Match 100.0%; Score 2238; DB 10; Length 2238;
Best Local Similarity 100.0%; Pred. No. 0;
.Matches 2238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGAATACAAACACTCAGATCTTCTACATATAACAAACACTGCGGAAATGC 60
|||||
DB 1 ATGGAATACAAACACTCAGATCTTCTACATATAACAAACACTGCGGAAATGC 60
QY 61 CCTTTTACCGAGGTTTCGCTTAAGCAAGTGCAGGTGGCGGACCAACAGGATTGG 120
|||||
DB 61 CCTTTTACCGAGGTTTCGCTTAAGCAAGTGCAGGTGGCGGACCAACAGGATTGG 120

QY	121	TGGCCCAACATGCTCAACCTCGGCATCTTAGCCCAACATTCATCGCTATCGGACCCAAAC	180
Db	121	TGGCCCAACATGCTCAACCTCGGCATCTTAGCCCAACATTCATCGCTATCGGACCCAAAC	180
QY	181	GACCCGGATTTGACTATGCCGAAGAGTTTAAAGAGCTAGATCTGGCAGCGGTTAAAGAAG	240
Db	181	GACCCGGATTTGACTATGCCGAAGAGTTTAAAGAGCTAGATCTGGCAGCGGTTAAAGAAG	240
QY	241	GACCTGGCAGCGCTAATGACAGATTCACAGAGCTGGTGGCCAGAGATTTACGGTCATTAT	300
Db	241	GACCTGGCAGCGCTAATGACAGATTCACAGAGCTGGTGGCCAGAGATTTACGGTCATTAT	300
QY	301	GGCCCTCTTTATACGATGGCGTGGCAGACGCGCGACCTACCGTATCGGTATGGC	360
Db	301	GGCCCTCTTTATACGATGGCGTGGCAGACGCGCGACCTACCGTATCGGTATGGC	360
QY	361	CGTGGTGGCGTGGCTCGGCTCACAGCGCTTCGCGCTCTCAATAGCTGGCCAGACAAAT	420
Db	361	CGTGGTGGCGTGGCTCGGCTCACAGCGCTTCGCGCTCTCAATAGCTGGCCAGACAAAT	420
QY	421	GCCAACTGGATAAAGCAGCGTTGCTTCTTTGGCCCATCAAAACAAATACGGTCGAAA	480
Db	421	GCCAACTGGATAAAGCAGCGTTGCTTCTTTGGCCCATCAAAACAAATACGGTCGAAA	480
QY	481	ATCTCTGGCGGATCTAATGATACTCACAGGAACGTAGCTCTGGAACTATGGGCTTT	540
Db	481	ATCTCTGGCGGATCTAATGATACTCACAGGAACGTAGCTCTGGAACTATGGGCTTT	540
QY	541	AAAACTTTTGGTTTGGAGTGGCAGAGCAGATGATGGGAGCCTGAAGAAGATGTATAC	600
Db	541	AAAACTTTTGGTTTGGAGTGGCAGAGCAGATGATGGGAGCCTGAAGAAGATGTATAC	600
QY	601	TGGGAGCAGAAACCGAATGGCTGGGAGACAAGCGCTATGAAGTGCACCGAGAGCTCGAA	660
Db	601	TGGGAGCAGAAACCGAATGGCTGGGAGACAAGCGCTATGAAGTGCACCGAGAGCTCGAA	660
QY	661	AATCCCTGGAGCGGTACAAATGGGACTCATCTATGTAACCCCGGAGGACCCAAAGCG	720
Db	661	AATCCCTGGAGCGGTACAAATGGGACTCATCTATGTAACCCCGGAGGACCCAAAGCG	720
QY	721	AAGCCAGACCCATCGCTGCTGCGGTGATATCGTGAGACTTTTGGCCGAATGGCAATG	780
Db	721	AAGCCAGACCCATCGCTGCTGCGGTGATATCGTGAGACTTTTGGCCGAATGGCAATG	780
QY	781	AATGACCAAGAACCGTGGCTCTCATAGCGGTGGACACACCTTTGGGAAACCCATGGT	840
Db	781	AATGACCAAGAACCGTGGCTCTCATAGCGGTGGACACACCTTTGGGAAACCCATGGT	840
QY	841	GCTGCCGATCGGAGAAATATGTGGCCGAGAGCGCTCGCGCGCAGGTATTGAAGAAATG	900
Db	841	GCTGCCGATCGGAGAAATATGTGGCCGAGAGCGCTCGCGCGCAGGTATTGAAGAAATG	900
QY	901	AGCCTGGGTGGAAAAACACTAGCGCACCGGACACCGTTCGGGATACCATCACCAGTGA	960
Db	901	AGCCTGGGTGGAAAAACACTAGCGCACCGGACACCGTTCGGGATACCATCACCAGTGA	960
QY	961	CTAGAAGCGCTGGACCAAGACCTAGCGCACCGGACACCGTTCGGGATACCATCACCAGTGA	1020
Db	961	CTAGAAGCGCTGGACCAAGACCTAGCGCACCGGACACCGTTCGGGATACCATCACCAGTGA	1020
QY	1021	TTTGGTTACAGTGGGAGCTTACCAAAAGTCCAGCTGGAGCTTATCAGTGGAAACCAAAA	1080
Db	1021	TTTGGTTACAGTGGGAGCTTACCAAAAGTCCAGCTGGAGCTTATCAGTGGAAACCAAAA	1080
QY	1081	GACGGTCCGGGTGGCACCATTACCGGATGCACATGATCCCGAAGTGGCAGCTTCCA	1140
Db	1081	GACGGTCCGGGTGGCACCATTACCGGATGCACATGATCCCGAAGTGGCAGCTTCCA	1140
QY	1141	TTTATGCTCACTACGGACCTGGCGCTCGCATGAGACCCGTGATTACGAAAAAATTTCTCGA	1200
Db	1141	TTTATGCTCACTACGGACCTGGCGCTCGCATGAGACCCGTGATTACGAAAAAATTTCTCGA	1200
QY	1201	CGGTACTGAAACCCCTGATGAGTTTGGAGATGCTTTCGCGAAGACNTGGTACAAACTG	1260

RESULT 2

Db	1201	CGGTACTGAAACCCCTGATGAGTTTGCAGATGCTTTCGGAAGAGCATGTGTACAAACTG	1260
QY	1261	ACACACAGAGATATGGGACCAAGTGCGCTACCTGGGACCAAGTGCCTTCAGGAAGAC	1320
Db	1261	ACACACAGAGATATGGGACCAAGTGCGCTACCTGGGACCAAGTGCCTTCAGGAAGAC	1320
QY	1321	CTCATCTGGCAAGACCCCTATACAGATGTAAGCCATCTCTTGTAGACGAAACGATATT	1380
Db	1321	CTCATCTGGCAAGACCCCTATACAGATGTAAGCCATCTCTTGTAGACGAAACGATATT	1380
QY	1381	GAAGGCTTAAAGCCAAATCTCGGACTGAGGTAAGCGGTAAGCGAGCTGGTAAGCAGC	1440
Db	1381	GAAGGCTTAAAGCCAAATCTCGGACTGAGGTAAGCGGTAAGCGAGCTGGTAAGCAGC	1440
QY	1441	GCATGGGCTTCTGCATCTACTTTAGAACTCTGACAAAGCGCGGTGCAACGCTGCA	1500
Db	1441	GCATGGGCTTCTGCATCTACTTTAGAACTCTGACAAAGCGCGGTGCAACGCTGCA	1500
QY	1501	CGTATACGACTGGCCCAACAAAAGACTGGGAAGTAAACACCTCAGCAACTTGCACAGG	1560
Db	1501	CGTATACGACTGGCCCAACAAAAGACTGGGAAGTAAACACCTCAGCAACTTGCACAGG	1560
QY	1561	GTACTCAAAACACTAGAAAGTATCCAGAGGACTTTAACCCAGGCGCAATCAGATAACAAA	1620
Db	1561	GTACTCAAAACACTAGAAAGTATCCAGAGGACTTTAACCCAGGCGCAATCAGATAACAAA	1620
QY	1621	GCATATCTGTTGGCGGACCTGATTTGCTGGCCGCTGTGGGTGTAGAAAAAGCTGCA	1680
Db	1621	GCATATCTGTTGGCGGACCTGATTTGCTGGCCGCTGTGGGTGTAGAAAAAGCTGCA	1680
QY	1681	AAAGATGCTGGCCATGAGTSCAGGTGCTTTTAAACCCGGGACAGCGGATGCCACCGCT	1740
Db	1681	AAAGATGCTGGCCATGAGTSCAGGTGCTTTTAAACCCGGGACAGCGGATGCCACCGCT	1740
QY	1741	GAGCAAAACCGTGTGGAAAGCTTTTCGAAGCACTAGAGCCAGCGGCTGAGGGCTTTAGAAAC	1800
Db	1741	GAGCAAAACCGTGTGGAAAGCTTTTCGAAGCACTAGAGCCAGCGGCTGAGGGCTTTAGAAAC	1800
QY	1801	TACATTTAAACCGGAGCATAAAGTATCGCTGAGCAAAATGCTCTGAGACCGGCGCAGCTT	1860
Db	1801	TACATTTAAACCGGAGCATAAAGTATCGCTGAGCAAAATGCTCTGAGACCGGCGCAGCTT	1860
QY	1861	CTGTGCTTTCGGGACCAAGAAATGACTGCTTTGGTAGCGGTATGCTGTACTGGGCAAC	1920
Db	1861	CTGTGCTTTCGGGACCAAGAAATGACTGCTTTGGTAGCGGTATGCTGTACTGGGCAAC	1920
QY	1921	AATGACGAGCTTGGCAGCATGGAGTGTACAAATTAAGCGGCTCAGCTATCCAAATGAC	1980
Db	1921	AATGACGAGCTTGGCAGCATGGAGTGTACAAATTAAGCGGCTCAGCTATCCAAATGAC	1980
QY	1981	TTCTTTTAAACCTGTAGACCTCAACACTAAATGGCGAGCCAGCGATGAATCAGACAAA	2040
Db	1981	TTCTTTTAAACCTGTAGACCTCAACACTAAATGGCGAGCCAGCGATGAATCAGACAAA	2040
QY	2041	GTTTTGAAGGACAGACTTCAAAACTGGGGAAGTAAAGTGGAGTGGCAGCCCGGGTAGAC	2100
Db	2041	GTTTTGAAGGACAGACTTCAAAACTGGGGAAGTAAAGTGGAGTGGCAGCCCGGGTAGAC	2100
QY	2101	CTGATCTTCGATCCCAATTCGAGCTAAGAGCCCTCCGAGNAGTGTACGCTGTGCAGAT	2160
Db	2101	CTGATCTTCGATCCCAATTCGAGCTAAGAGCCCTCCGAGNAGTGTACGCTGTGCAGAT	2160
QY	2161	TCGGAAGAAAGTTTGTAAAGATTTTGTGAAGCCCTGGGCGCAAGTAAATGGACCTGGAC	2220
Db	2161	TCGGAAGAAAGTTTGTAAAGATTTTGTGAAGCCCTGGGCGCAAGTAAATGGACCTGGAC	2220
QY	2221	CGGTTTGTATCTGAAATAA 2238	
Db	2221	CGGTTTGTATCTGAAATAA 2238	

US-10-080-170-645/c
; Sequence 645, Application US/10080170
; Publication No. US20030129601A1
; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
; IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 645
; LENGTH: 82993
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-080-170-645

Query Match 37.7%; Score 844.4; DB 15; Length 82993;
Best Local Similarity 63.0%; Pred. No. 8.4e-265;
Matches 1390; Conservative 0; Mismatches 801; Indels 15; Gaps 5;
QY 30 TAGGTATACACAACTGGCGGAAATGCCCTTTTACGGAGGTTGCTTAAGCAAAG 89
DB 40475 TACAGAAACCAACACCGGAGCGCTAGCAACGGCTGTCCCGTGTGGGTGCATATGAATA 40416
QY 90 TGCAGTGGCGGACCAACAAAGGATTGGTGGCCCAACATGCTCAACCTCGGCATCTT 149
DB 40415 CCCGTGCGAGGGCGGGAACCGAGCTGGTGGCCCAACCGCTCAATCTGAGGTACT 40356
QY 150 ACGCCAACAATCATCGCTATCGGACCCAAACAGCCGATTTTGACTATGCCAAGAGTT 209
DB 40355 GCACCAAAACCCCGCGTGTGCTGACCCGATGGTGGCGGTTCGACTATGCCCGGAGGT 40296
QY 210 TAGNAGCTAGATCTGCGACGGTTTAAAGGACCTGGCAGCGCTAATGACAGATTACA 269
DB 40295 CGGCACCATCGAGCTTGACGCCCTGACGGGGACATCGAGGAAGTATGACCACTCGCA 40236
QY 270 GGAAGTGGCCAGCAGATTACGGTTCATTATGGCCCTCTTTTATACGATGCGGTGGCA 329
DB 40235 GCGGTGGTGGCCGCGCAGCTACGGCCACTACGGGCGCGCTTTATCCGATGCGGTGGCA 40176
QY 330 CAGCGCGGACCTACCGTATCGGTGATGCGCTGTGGCGGTGGTCCGGCTCACAGCG 389
DB 40175 CGTGGCGGACCTACCGCATCCACGACGCGCGCGCGCGCGGCGGCGCATGCGCG 40116
QY 390 CTTGCGGCTCTCAATAGCTGGCCAGCATGCAATCTGGATAAGCAGCTTGTCTTCT 449
DB 40115 GTTCGGCGCTTAAAGCTTGGCCCGGACACGCGCAGCTTGGACAGCGCGCGCGGTGCT 40056
QY 450 TTGGCCCATCAACAAATAATACGGTGGAAATAATCTCTGGCGGATCTTAATGATCTCAC 509
DB 40055 GTGGCGGTCAAGANGATGACGCAAGAGCTCTCATGGCGGACCTGATTTTTCG 39996
QY 510 AGAAACGTAGCTCTGGAACATATGGCTTTAAACATTTTGGTTTTCAGGTGGCAGAGC 569
DB 39995 CGGCAACTGCGCTGGAATCGATGGCTTCAAGAGCTTCGGGTTTCGGCTTCGGCGGGT 39936
QY 570 AGATGTTATGGAGCTTGAAGAGATGATATCTGGGAGCAGAAACCGAATGGCTGGAGA 629
DB 39935 CGACCAAGTGGAGGCC---CGATAGGTCTATTGGGCAAGGAAGCAACCTGGCTCGGCA 39879
QY 630 CAAGCGCTATGAAGTGAACGAGAGCTCGAAATAATCCCTGGGAGCGGTACAAATGGACT 689
DB 39878 TGAGCGTTACAGCGGTAAAGCGGATCTGGAGAACCCGCTGGCGCGCGGTTCAGATGGGGT 39819
QY 690 CATCTATGAACCCCGAAGGACCCAAACCGCAAGCCAGACCTTATCGCTGCTCGCGGTGA 749
DB 39818 GATCTACGTGAACCCCGAGGGGCGGAACGCGACCCCGGACCCCATGCGCGCGCGGTGCA 39759

QY 750 TATTGCTGAGACTTTTGGCCGAATGCAATGACGAAGAAACCGTGGCTTCTCATAGC 809
DB 39758 CATTCGCGAGAGCTTTTCGGCGCATGCCATGACGACGTCGAACAGCGCGCTGATCGT 39699
QY 810 GGTGTGACACACCTTCGGAAACCCATGCTGCTGCCGATGCCGAGAAATATATGGGCGC 869
DB 39698 CGCGGTGCACACTTCGGTAAGACCCCATGCGC---CGGCCGCGGCTGCTGTCGGCCC 39642
QY 870 AGAGCCTGCGCGCAGAGTATTGAAGAAATGAGCCTGGGTGGGTAAGAAACACCTTACGGCAC 929
DB 39641 CGAACCCGAGGCTGCTCCGCTGAGCAGATGGCTTTGGGTGGAAGAGCTCGTATGGCAC 39582
QY 930 CGGACACGGTGGGATACCATCAGTGGACTAGAAAGCGCTGGACAAAGACCCCTAC 989
DB 39581 CGGAACCGTAAGGACGCGATCACAGCGCATCGAGTGTGTGACCAACACCCCGAC 39522
QY 990 TCAATGAGCAATAACTTTTTCGAAACCTTTTGGTTAGAGTTCAGAGTGGAGCTGACGAAG 1049
DB 39521 GAATGGGACACAGTTTCTCGAGATCTGTACGGCTACGAGTGGAGCTGACGAAGAG 39462
QY 1050 TCCAGCTGGAGCTTATCAGTGGAAACCAAAAGACGCTGGCGGGCTGGCACCATACCGGA 1109
DB 39461 CCTGCTGGCGCTTGGCAATACACCCCAAGGACGCGCGGTGGCGCACCATCCCGGA 39402
QY 1110 TGCACATGATCCAGCAAGTCGACGCTCATTTATGCTCAGTACGAGACCTGGCGGTGCG 1169
DB 39401 CCGGTTTCG---GGGGCCAGGGCGCTCCCGAGCATGCTGGCACATGACCTCTCGGTGCG 39345
QY 1170 CATGGACCTGATTACGAAATAATTTCTGACCGGTACTATGAAACCCCTGATGATGCTGC 1229
DB 39344 GGTGGATCGATCTATGAGCGGATCAGCGTCTGCTGGTGGAAACACCCCGAGGANTGGC 39285
QY 1230 AGATGCTTTCGAAAGCATGTTACAACTGACACACAGAGATATGGGACCAAGGTGCG 1289
DB 39284 CGACGAGTTCCGCAAGCGCTGTTACAAGTGTATCCACGAGACATGGTCCCGCTTCGAG 39225
QY 1290 CTACCTGGGACCAAGAGTGCCTCAGGAGACCTCATCTGGCAGAGACCTTATACCAATGT 1349
DB 39224 ATACCTTGGGCGCTGCTCCCCAAGCAGACCCCTGCTGTGGCAGGANTCCGGTCCCTCGGT 39165
QY 1350 AAGCCATCTCTTGTAGACGAAACGATATTGAAGCCTTAAAGCCTTAAAGCCTTGAATC 1409
DB 39164 CAGCCAGACCTGCTGGCGGACCGGAGATGCGACGCTTAAAGCAGCATCCGGGATC 39105
QY 1410 GGAAGTACGAGTAAAGCAGCTGTTAAGCAGCGGCTTCTGCTATCTACTTTTATGAAA 1469
DB 39104 GGAATGACTGTCTCAGCTAGTTTCGACCGCATGGCGGCGGCTGCTGCTGCTGCTG 39045
QY 1470 CTCTGACAGCGGCGGTGCCAACGTCACGATACGACTGGCCCGCCACAAAGACTG 1529
DB 39044 TAGCGAAGCGGCGGCGCCAACGTTGGTCCGCTCCGCTCGACCCACAACTCGGGT 39985
QY 1530 GGAAGTAAACAAACCCCTCA---GCAACTTGGCAGGCTACTCAAAACACTAGAAGTATCCA 1586
DB 39984 GGAGGTCAACGACCCCGACGGGATCTGCGCAAGGTCATTCGACCTTGAAGAGATCCA 39925
QY 1587 GGAGGACTTTAACAGCGCAATCAGATACAAAGCAGTATCGTTGGCGACCTGATGTT 1646
DB 39924 GGAGTCATTCAACTCCGCGCGCGCGGGAACATCAAAAGTGTCTCTCCCGACCTCGCT 38865
QY 1647 GCTGGCGGCTGTGCGGGTGTAGAAAAGCTGCAAAAGATGCTGGCCATGAGTGCAGGT 1706
DB 39864 GCTCGTGGCTGTGCGGCCATAGAAAAGCAGAAAGCGGCGCTGGCCACAACTACAGGT 39805
QY 1707 GCCTTTCAACCCCGGACGAGCGGATGCCACGCTGAGCAACCGATGTGGAAGCTTTCGA 1766
DB 39804 GCCTTACCCCGCGCGCGGATGCGTGGAGAACAAACCGACGAGTGGAAATCTTTGC 39745
QY 1767 AGCACTTAGGCCAGCGGCTGACGGCTTTAGAACTACTATTAACCCCGAGGATAAGTATC 1826
DB 39744 CGTGTGGAGCCCAAGGACAGTGGCTTCGAAACTACCTCGGAAAGGGCAACCCGTTGCC 39685
QY 1827 CGCTAGGAAATGCTCGTAGACCGGCGCAGCTTCTGTGCTTTTCGGCACCAAGAAATGAC 1886

```

Db      38684  GCGCCAGTACATGCTGCTCGACAGCGCAACCTGCTTACGCTCAGTGCCCTCGAATGAC 38625
QY      1887  TCGTTTGTAGCGGTATGCGGTACTGGGACCAACCTACGACGGTTCGCCAGCATGGAGT 1946
Db      38624  GGTGCTGTAGTGGCTCGCGTCTCGCGCAAACTACAAGCGCTTACCGCTGGCGT 38565
QY      1947  GTTACAAATAAGCGGGTACGATATCAATGACTCTCTTTGTAAACCTGTAGACCTCAA 2006
Db      38564  GTTACCGAGCGCTCCGAGTCACTGACCAAGCACTTCTGTGAACCTGTCTCGACATGGG 38505
QY      2007  CACTAAATGGCGAGCGAGCATGAATCAGACAAAGTTTTTTGAAGCGCAGAGACTTCAAAAC 2066
Db      38504  TATCACCTGGGAGCGCTCGCCAGCAGATGACGGGACCTACAGGCAAGATGGC---AG 38448
QY      2067  TGGCGAAGTAAAGTGGAGTGGCACCCTGGGTAGACTGATCTTCGATCCAAATTCGAGCT 2126
Db      38447  TGGCAAGTGAAGTGGAGCGCGCAGCGGTGAGCTGTCTTCGGTCCAACTCGGAGTT 38388
QY      2127  AAGAGCCCTCCAGAGGTGACGGCTGTGCAAGATCTCGAAGAAAGTTTGTAAAGATT 2186
Db      38387  CGCGGCGCTTCTGAGGTCTATGCGCGCGATGACCGCGAGCGGAAGTTCTGTGAGGACT 38328
QY      2187  TGTGAAGCGCTGGCGCAAGTAATGGACCTGACCGGTTTGATCTG 2232
Db      38327  CGTGCCTGCCCTGGGCAAGGTGATGAACCTCGACAGGTTTCGACGTG 38282

```

RESULT 3

```

; US-09-884-889-5
; Sequence 5, Application US/09884889
; Patent No. US20020102680A1
; GENERAL INFORMATION:

```

```

; APPLICANT: DIVERSA CORPORATION
; APPLICANT: ROBERTSON, Dan
; APPLICANT: SANYAL, Indrajit
; APPLICANT: ADHIKARI, Robert
; TITLE OF INVENTION: CATALASES
; FILE REFERENCE: DIVER1100-4
; CURRENT APPLICATION NUMBER: US/09/884,889
; CURRENT FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 09/412,347
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: US 08/951,844
; PRIOR FILING DATE: 1997-10-16
; PRIOR APPLICATION NUMBER: US 08/674,887
; PRIOR FILING DATE: 1996-07-03
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5

```

```

; LENGTH: 2262

```

```

; TYPE: DNA
; ORGANISM: Alcaligenes (beleya) aquamarinus
; US-09-884-889-5

```

```

Query Match      30.08; Score 672.2; DB 10; Length 2262;
Best Local Similarity 59.5%; Pred. No. 3.2e+209;
Matches 1277; Conservative 0; Mismatches 808; Indels 60; Gaps 6;

```

```

QY      99  CGGCACCAAAACAGGAGTTGGTGGCCCAACATGCTCAACCTCGGCATCTTACGCCAACA 158
Db      150  CGGTACTTCCACAAAGATTGGTGGCGGAGGGTTGAACCTGGATTTTGATCAGCA 209
QY      159  TTCATCGCTATCGGACCCCAACGCCGATTTTGACTATCGCGAAGAGTTTAAAGAGCT 218
Db      210  AGATCGCAATCAGACCCGATGGATCGGATTTCAACTACCTGGAAGAAGTACGACGCT 269
QY      219  AGATCTGGCAGCGGTAAAGAGACCTGGCAGCGCTTAATCAGAGATTCACAGACTGGTG 278
Db      270  CGATTTCGAGCGCTGAAGAAGATGTCCACGCGTTGATACCGCATAGCAAGAGTGGTG 329
QY      279  GCCAGCAGATTACGGTCAATTATGGCCCTTCTTTATACGATGCGGTGGCAGACGCCCG 338

```

..

```

Db      330  GCGCCCTGACTGGGGGCACTAGGCGGTTTGATGATCCGTATGGCTTGGCATCCGCTGG 389
QY      339  CACCTACCGTATCGGTGATGCGCGTGGTGGGCTCCGCTCACAGCGCTTCGCGCC 398
Db      390  CACCTACCGTATGCTGATGCGCGTGGGCGGTTACCGGAAGCAGCGCTTTGCACC 449
QY      399  TCTCAATAGCTGGCCAGACAATGCCAATCTGGATAAAGCAGCTTTGCTTCTTTGGCCCAT 458
Db      450  GCTCAACTCTGGCGGCAACAGTCAAGCTGGATAAAGCGCGCGCTCTGCTGTGGCCGAT 509
QY      459  CAACAAAAATACGGTTCGAAAAATCTCTGGCGGATCTAATGATACCTACACAGGAACGT 518
Db      510  CAAGAAGAAGTACGGCAACAAATCAGCTGGGCAAGCTGATGATTTCTTGGCTGGCACGCT 569
QY      519  AGCTCTGAAACTATGGGCTTTTAAACCTTTTGGTTTTTTCAGGTGGCAGAGCAGATGATG 578
Db      570  GGCATTAGCTCCATGGCTTACCTGCTTACGGCTTCTCTTTCGCGCGCTCGATATTTG 629
QY      579  GGAGCTGAAGAAGATGATATCTGGGGAGCAGAAACCGAATGGCTGGGA-----GACAA 632
Db      630  GGAACCCGAAAAAGATATCTACTGGGTGACGAAAAAGAGTGGCTGGCACCTTCTGACGA 689
QY      633  GCGCTATGAAGGTGACCGAGAGC-----TGGAAAAATCCCTGGGAGCGGTACAAAT 683
Db      690  ACGCTACGGCGACGTGAACAAAGCCAGAGACCATGGAACCCGCTGGCGGCTGTCCAAAT 749
QY      684  GGGACTCATCTATGTAAACCCCGAAGGACCAAGCGCAAGCAGACCCCTATCGCTGTGTC 743
Db      750  GGGTCTGATCTATGTGAACCCGGAAGGTGTTAAGCGGCACCTGATCGCTTGAAACCGC 809
QY      744  GCGTGATATTCGTGAGACTTTTGGCCGAATGGCAATGAATGACGAAGAAACCGTGGCTCT 803
Db      810  ACAGCAGGTACTTGAACCTTTCGCCCGTATGGCGATGAACGACGCAAAAAACCGACGCT 869
QY      804  CATAGCGGTGGACACACCTTCGGAACCAACCCATGGTCTGCCCATGCGGAGAAATATGT 863
Db      870  CACAGCTGGCGGCCACACCGTCCGTAATGTACGCGTAATGGCAATTCGCTTGGCT---T 926
QY      864  GGGCCGAGAGCTGCCCGCGCAGGTATTGAAGAAATGAGCTGGGTGGGAAACACACTA 923
Db      927  AGCCCTTGACCCAAAGCCTCTGACGTTGAAACACGAGGCTTAGTTGGGCAACCCCA 986
QY      924  CGGCACCGGACACGGTGGGATACCATCAGCTGGAGTGAAGCGCGCTGGAGCAAGAC 983
Db      987  CATCGAGGCAAGGCAAGCAACGCCGTGACCTCGGGTATCGAAGGTGCTTGGACCAACAA 1046
QY      984  CCTTACTCAATGGAGCAATAACTTTTGAACACCTCTTGGTTACGAGTGGGAGCTTAC 1043
Db      1047  CCCACGAAATTCGATATGGGCTATTCGACCTCTGCTCGGCTACAAATTTGGGAACGAA 1106
QY      1044  CAAAGTCCAGCTGGAGCTTATCAGTGGAAACCAAGAGAGGTCGCCGGGCTGGCACCAT 1103
Db      1107  AAGAGTCTTGGCGTGCACCACCATTTGGAAACCGATTGACATCAAAAGGAACCAAGCC 1166
QY      1104  ACGGATGCACATGATCCGAGCAAGTCGACGCTCCATTTATGCTACTACGGACCTGGC 1163
Db      1167  GGTTCGCGCAGGACGCCCTCTATTTCGCCACAACCCGATCATACCGCATCGGATATGGC 1226
QY      1164  GCTCGCATGGACCTGATTAACGAAAAAATTTCTCGACGGTACTATGAAACCCCTGATGA 1223
Db      1227  GATAAAGGTAATTCGACCTATCGCGCTATCTGCGAAAAAATTCATGCGCGATCTCTAGTA 1286
QY      1224  GTTTGCAAGTCTTTCGGAAGCATGTTACAAACTGACACACAGAGATATGGACCAAA 1283
Db      1287  CTTCAAGAAACCTTCGGAAGGGTGTTCAGCTGACCGACCGTACCTGGCGCCGAA 1346
QY      1284  GTTGGCTACTGGGACCAAGTGTCTCAGGAAGCCTCATCTTGGCAAGACCTATFACC 1343
Db      1347  ATCAGCTTACATGGCGCCGGAAGTGGCGGCAAGAGACCTGATTTGGCAAGACCGGATTC 1406
QY      1344  AGATGAAGCCATCTCTTGTAGCAGAAACGATATTGAAGGCTTAAAGCCCAAAATCCT 1403
Db      1407  GCGAGGTAAC-----ACCGACTACTCGGAAGAGTGTCAAGCAGAAATTTGC 1454

```

```

QY 1404 GGAATCGGACGCTAGCGGAGAGCTGGTAAGCAGCGCATGGCTTCTGCACTACTTT 1463
Db 1455 ACAAAAGTGGCTGAGCATTAGTAGATGCTCCACCGCTTGGGACAGTGGCCGCTACTTA 1514
QY 1464 TAGAAACTCTGACAAAGCGGGGTGCCAACGGTGCAGTATACGACTGGCCCAACAAA 1523
Db 1515 TCGCGGTTCCGATATGGCGGGGGTGTAAACGGTGCAGCTGCTTGGCCCAACAGAA 1574
QY 1524 AGACTGGGAAGTAACAAACCCCTAGCAACTGCCAGGGTACTCAAAACACTAGAAAGTAT 1583
Db 1575 CGAGTGGAGGCAACAGCGGAGCGCTGCGGAAAGTGTCT----- 1616
QY 1584 CCAGGAGGACTTTAACCGGCGCAATCAGATACAAAGCAGTATCGTTGGCGGACCTGAT 1643
Db 1617 -----GAGCGTCTACGAGCAGATCTCTCCGACACCGCGGTAGCATGCGGAGGTGAT 1670
QY 1644 TGTGCTGGCGGCTGTGCGGGGTGTAGAAAAGCTGCAAAAGATGCTGGCCATGAGGTGCA 1703
Db 1671 CGTTCTGGCGGTAGCTAGGATCGAGAAAGCCGGAAGCAGCAGGTTCAGATGTGG 1730
QY 1704 GGTGCTTTCAACCCGGGACGAGCGGATGCCACCGCTGAGCAACCGATGTGGAAGCTTT 1763
Db 1731 CGTTCCCTTCTGAAAGCGGTGGCGATGCGACCGCGGAGATGACCGCAGACTCCTT 1790
QY 1764 CGAAGCACTAGACCGGCTGACGGCTTTAGAACTACATTAACCGGAGCATAAAGT 1823
Db 1791 CGACCGCTGGAGCGCGCTGGCGGATGGCTTCGCGCACTGGCAGAAAGAGATGTGT 1850
QY 1824 ATCCGCTGAGGAAATGCTCGTAGACCGGGCGCAGCTTCTGTGCTTTCGGCACCAAGAT 1883
Db 1851 GAAGCGGAGAGATGCTCTGATCGTGGCAGCTGATGGCTTAACCGGCCGGAAT 1910
QY 1884 GACTGCTTTGAGCGGTATCGGTGTAAGTGGGACCACTACGACGGTTCGACGATGG 1943
Db 1911 GACCGTGTCTGGGCGGTATGCGGTACTGGCACCACTATGTTGGCACCAACACCG 1970
QY 1944 AGTGTTTACAATAAGCGGCTGAGTATCCATGACTTCTTTGTAACCTGCTAGACCT 2003
Db 1971 CGTATTCACGATGTGAAGGCGAGTGTGACCAACGACTTTTGTGAACCTGACCGATAT 2030
QY 2004 CAACACTAAATGGCGAGCGCGATGAATCAGACAAAGTTTTGAAGCGAGAGACTTCAA 2063
Db 2031 GGGGACACTGG-----AAGCGGTAGGTAGCAACGCTACGAAATCCGCGACCGCAA 2084
QY 2064 AACTGGCAAGTAAAGTGGAGTGGCACCGGGTAGACCTGATCTTCGGATCCAATTCGA 2123
Db 2085 GACCGTGGCTGAAAGTGGACCGCTCGCGGTGGATCTGTTGTTGTTCAACTCGCT 2144
QY 2124 GCTAAGAGCCCTCGCAGAGTGTACGGCTGTGCAATCTGAAGAAAAGTTTGTAAAGA 2183
Db 2145 ACTGCGCTCTTACGAGAAAGTGTACGCCAGGACGATTAACCGGAGAGTTCGTACAGA 2204
QY 2184 TTTTGTGAGGCGCTGGGCCAAAGTAAATGACCTGGACCGGTTGA 2228
Db 2205 CTGCTGCGCGCTGGACCAAGTATGAACCGCCGACCGCTTTCGA 2249

```

RESULT 4

US-09-940-925A-145
 : Sequence 145, Application US/09940925A
 : Publication No. US20030054338A1
 : GENERAL INFORMATION:

APPLICANT: BROW, MARY ANN D.
 LYANICHEV, VICTOR I.

OLIVE, DAVID M.

TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF

PATHOGENS

NUMBER OF SEQUENCES: 165

CORRESPONDENCE ADDRESS:

ADDRESS: MEDLEN & CARROLL

STREET: 220 MONTGOMERY STREET, SUITE 2200

CITY: SAN FRANCISCO

:

STATE: CALIFORNIA
 COUNTRY: UNITED STATES OF AMERICA
 ZIP: 94104
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/940,925A
 FILING DATE: 10-Jun-2002
 CLASSIFICATION: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: CARROLL, PETER G.
 REGISTRATION NUMBER: 32,837
 REFERENCE/DOCKET NUMBER: FORS-01756
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 705-8410
 TELEFAX: (415) 397-8338
 INFORMATION FOR SEQ ID NO: 145:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 620 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 SEQUENCE DESCRIPTION: SEQ ID NO: 145:
 US-09-940-925A-145

```

Query Match      11.3%; Score 251.8; DB 11; Length 620;
Best Local Similarity 63.9%; Pred. No. 1.6e-71;
Matches 398; Conservative 0; Mismatches 22; Indels 3; Gaps 1;

QY 916 AACACCTAGCGGACCGGACACGCTGGGATACCATCACCAGTGGACTAGAAAGCGCGTGG 975
Db 1 AGCTCGATGGCACCGGACCGGTAGGACGCGATCACCAGCGCATCGAGTCTGATGG 60
QY 976 ACCAAGACCCCTACTCAATGGAGCAATACTTTTGAACAACTCTTTGTTACGAGTGG 1035
Db 61 ACGAACACCCCGACGAAATGGGACAAACAGTTTCTCGAGATCTCTACGCTACGAGTGG 120
QY 1036 GAGCTTACCAAAAGTCCAGCTGGAGCTTATCAGTGGAAACCAACCAACAGCGTGGCGGGCT 1095
Db 121 GAGCTGACGAAGAGCCCTGTGCTGGCTTGGCAATACACCGCCAAAGACGCGCGCTGGCC 180
QY 1096 GGCACCATACCGGATGCACATGATCCAGCAGTCCAGCTCCATTTATGCTCACTACG 1155
Db 181 GGCACCATCCCGGACCCGCTTGC---CGGGCCAGGGCGCTCCCGGACGATGCTGGCCACT 237
QY 1156 GACCTGGCGCTGGCGATGAGACCTGATTACGAAAAAATTTCTCGACGGTACTATGAAAC 1215
Db 238 GACCTCTCGCTGGCGGTGGATCGGATCTATGAGCGGATCAGCGGTGCGCTGGCTGGAACAC 297
QY 1216 CTTGATGAGTTTGCAGATGCTTTCCGGAAGCATGTGTACAACTGACACACAGAGATATG 1275
Db 298 CCGGAGAAATTTGGCCGACGAGTTCGCCAAGGCGCTGTGTACAACTGATCACCAGGACATG 357
QY 1276 GGACCAAAAGTGGCGTACCTGGGACCAAGTGCCTCAGGAAGCTCATCTCGCAAGAC 1335
Db 358 GGTCCGTTGGCAGATACCTTGGCGCGCTGGTCCCAAGCAGACCTGCTGTGGCAGGAT 417
QY 1336 CCTATACAGATGTAAAGCCATCCTTTGTAGACGAAAAACGATATGAAGGCCCTAAAAGCC 1395
Db 418 CCGGTCCCTGGGTGACGCCACGACCTCGTCCGGGAAGCCGAGATTCCAGCCCTTAAGAGC 477
QY 1396 AAATCTCGAATCGGACTGACGGTAAAGGAGCTGGTAAGCAGCGCATGGGCTTCTGCA 1455
Db 478 CAGATCTCGCATCGGGATTGACTGTCTCACAGCTAGTTTTCGACCCGATGGCGGCGG 537
QY 1456 TCTACTTTTGAAGAACTCTGACAAAGCGCGGCTGGCCAAAGCTGACGATACGACTGGCC 1515
Db 538 TCGTCTGCTCCGTGGTAGCAGACAGCGCGGCGGCAACGGTGGTCTCGCATCCGCGCTGCAG 597

```

QY 1516 CCACAAAAGACTGGGAAGTAAA 1538
 Db 598 CCACAAGTCGGGTGGGAGGTCAA 620

RESULT 5

US-09-940-925A-149/c
 ; Sequence 149, Application US/09940925A
 ; Publication No. US20030054338A1
 ; GENERAL INFORMATION:
 ; APPLICANT: BROW, MARY ANN D.
 ; LYAMICHEV, VICTOR I.
 ; OLIVE, DAVID M.
 ; TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
 ; PATHOGENS
 ; NUMBER OF SEQUENCES: 165
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MEDLEN & CARROLL
 ; STREET: 220 MONTGOMERY STREET, SUITE 2200
 ; CITY: SAN FRANCISCO
 ; STATE: CALIFORNIA
 ; COUNTRY: UNITED STATES OF AMERICA
 ; ZIP: 94104
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/940,925A
 ; FILING DATE: 10-Jun-2002
 ; CLASSIFICATION: <unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: CARROLL, PETER G.
 ; REGISTRATION NUMBER: 32,837
 ; REFERENCE/DOCKET NUMBER: FORS-01756
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 705-8410
 ; TELEFAX: (415) 397-8338
 ; INFORMATION FOR SEQ ID NO: 149:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 620 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 149:

US-09-940-925A-149
 Query Match 11.3%; Score 251.8; DB 11; Length 620;
 Best Local Similarity 63.9%; Pred. No. 1.6e-71;
 Matches 398; Conservative 0; Mismatches 222; Indels 3; Gaps 1;
 QY 916 AACACCTACGCGACCGGATACCATCACCAGTGGAGTGGAGCGGCTGG 975
 Db 620 AGTCGTGTGGACCGGACCGGTAAGGACGCGATCACCAGCGCATCGAGGTG 561
 QY 976 ACCAAGACCCCTACTCAATGAGCAATACTTTTGAACACCTCTTTGGTTACGAGTGG 1035
 Db 560 ACGAACACCCGACGAAATGGACACAGTTTCTCGAGATCTGTACGCTACGAGTGG 501
 QY 1036 GAGCTTACCAAAAGTCCAGCTGAGCTTATCAGTGGAAACCAAAAGACGCGCGGGCT 1095
 Db 500 GAGCTGACGAAGACCCCTGTGGCGTTGGCAATACACCGCAAGGACGCGCGGTGCC 441
 QY 1096 GGACACCATCCGGATGACATCCGACGAAGTCGACCGCTCCATTTATGCTCACTACG 1155
 Db 440 GGCACCATCCCGACCCGTTGCG--GCGGCCAGGGCGGCTCCCGACGATGCGGCAC 384
 QY 1156 GACCTGGCGTGGCGATGGACCTGATTACGAAAAAATTTCTCGACGGTACTATGAAAC 1215
 Db 383 GACCTCTCGCTGCGGGTGGATCCGATCTATGAGCGGATCAGCGTCCGCTGGCTGGACAC 324

... ..

QY 1216 CCTGATGAGTTTGCAGATGCTTTCCGAAAGCATGGTACAACTGCACACAGAGATATG 1275
 Db 323 CCCGAGGAATTGGCCGAGAGTTTCGCCAAGGCCTGGTACAGCTGTATCCACCGAGACATG 264
 QY 1276 GGACCAAGGTGCGCTACCTACCTGGGACCAAGAGTGCCTCAGGAAGACCTCATCTGCAAGAC 1335
 Db 263 GGTCCCGTTGCGAGATACCTTTGGCCCGCTGCTGCCAAGCAGACACCTGCTGTGCAGAT 204
 QY 1336 CCTATACAGATGTAAGCCATCCTTGTGTAGACCAAAACGATATTGAAGGCTTAAAGCC 1395
 Db 203 CCGTCCCTGGGTGACGACGACCTCGTCGGCAAGCCGAGATTGCCAGCTTTAAGAGC 144
 QY 1396 AAAATCCTGGAATCGGACTGACGGTAAAGCAGCTGGTAAACGCGCATGGGCTTCTGCA 1455
 Db 143 CAGATCCTGCATCGGATTGACTGTCTCACAGCTAGTTTCGACCGCATGGCGCGCGG 84
 QY 1456 TCTACTTTTGAAGAACTCTGACAAGCGGGGGTGCACAGCGTGCACGTATACGACTGGCC 1515
 Db 83 TCGTCGTTCCGTGGTAGCACAAGCGGGCGGCCCAACGGTGTGCGATCCGCTCGAC 24
 QY 1516 CCACAAAAGACTGGGAAGTAAA 1538
 Db 23 CCACAAGTCGGGTGGGAGGTCAA 1

RESULT 6

US-09-941-193A-145
 ; Sequence 145, Application US/09941193A
 ; Publication No. US20030108873A1
 ; GENERAL INFORMATION:
 ; APPLICANT: BROW, MARY ANN D.
 ; LYAMICHEV, VICTOR I.
 ; OLIVE, DAVID M.
 ; TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
 ; PATHOGENS
 ; NUMBER OF SEQUENCES: 165
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MEDLEN & CARROLL
 ; STREET: 220 MONTGOMERY STREET, SUITE 2200
 ; CITY: SAN FRANCISCO
 ; STATE: CALIFORNIA
 ; COUNTRY: UNITED STATES OF AMERICA
 ; ZIP: 94104
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/941,193A
 ; FILING DATE: 28-Aug-2001
 ; CLASSIFICATION: <unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: CARROLL, PETER G.
 ; REGISTRATION NUMBER: 32,837
 ; REFERENCE/DOCKET NUMBER: FORS-01756
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 705-8410
 ; TELEFAX: (415) 397-8338
 ; INFORMATION FOR SEQ ID NO: 145:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 620 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 145:
 US-09-941-193A-145

Query Match 11.3%; Score 251.8; DB 11; Length 620;
 Best Local Similarity 63.9%; Pred. No. 1.6e-71;
 Matches 398; Conservative 0; Mismatches 222; Indels 3; Gaps 1;


```

; TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
; PATHOGENS
; NUMBER OF SEQUENCES: 165
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/940,925A
; FILING DATE: 10-Jun-2002
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: FORS-01756
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 143:
; SEQUENCE CHARACTERISTICS:
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 143:
US-09-940-925A-143

```

```

Query Match      11.2%; Score 250.2; DB 11; Length 620;
Best Local Similarity 63.7%; Pred. No. 5.4e-71;
Matches 397; Conservative 0; Mismatches 223; Indels 3; Gaps 1;

QY 916 AACACCTAGCGACCGGACGCGGATACCATCAGGTGGGATAGAGCGGCTGG 975
Db 1 AGCTCGTATGGACCGGACCGGATAGAGCGGATACCATCAGGTGGGATAGAGCGGCTGG 60

QY 976 ACCAAGACCCCTACTCAATGGAGCAATACTTTTGAACCACTCTTTGTTACGAGTGG 1035
Db 61 ACGAACACCCGAGGAATGGGACACAGTTTCTCGAGATCTCTGACGCTACGAGTGG 120

QY 1036 GAGCTTACCAAAAGTCCAGCTGGAGCTTATCAGTGGAAACCAAGACGCGGCGGCT 1095
Db 121 GAGCTGACGAAGAGCGCTGCTGGCGCTTGGCAATACACCGCCAAAGGACGCGCGGTGC 180

QY 1096 GGCACCATACCGGATGCACATGATCCAGCAAGTCGACGCTCCATTTATGCTCATAG 1155
Db 181 GGCACCATACCGGATGCACATGATCCAGCAAGTCGACGCTCCATTTATGCTCATAG 237

QY 1156 GACCTGGCGCTGGCATGACCCCTGATTTACGAAAAAATTTCTGACGGTACTATGAAAC 1215
Db 238 GACCTTCGCTGGGCTGGATCGGATCTATGACGGGATCAGCGTCCGCTGGCTGGAAAC 297

QY 1216 CCTGATGATTTGACAGTCTTTTCGCGAAAGCATGTTACAACTGACACACAGATATG 1275
Db 298 CCGGAGGAATGGCCGACGAGTTCGCAAGGCTGGTACAAAGTGTATCCACGAGACATG 357

QY 1276 GACCAAAAGTGGCTTACCTGGGACGAGAGTGCCTCAGAGAGACCTCATCTGGGAAGC 1335
Db 358 GGTTCGCTGGGATACCTTGGGCGCTGGTCCCAAGCAGACACCTGCTGTGGCAGGAT 417

QY 1336 CCTATACAGAGTGAAGGATCCTCTTGTAGAGCAAAAGCATTTAAAGGCTTAAAGGC 1395
Db 418 CCGGTCCTTCGCTGACGACACGACCTCTGTCGGGAAGCGGAGNTGCCAGCCTTAGAGC 477

QY 1396 AAAATCCTGGAATCGGGACTGACGGTAAAGGAGTGGTAAAGCAGCGCATGGGCTTCTGCA 1455

```

```

Db 478 CAGATCCGGCATCGGATTTGACTGTCTCACAGCTAGTTTCGACCGCATGGCGGGCGG 537
QY 1456 TCTACTTTTGAACACTCTGACAAGCGCGGCGGTCGCAACGGTGCAGCTATACGCTGGCC 1515
Db 538 TCCTCGTTCGTTGAGTGGTACGCAAGCGGCGGCGGCGGCAACGGTGTGCTGCGATCCGCTGCAG 597
QY 1516 CCACAAAAGACTGGGAAGTAAA 1538
Db 598 CCACAAAGTCGGTGGGAGGTCAA 620

RESULT 9
US-09-940-925A-146
; Sequence 146, Application US/09940925A
; Publication No. US20030054338A1
; GENERAL INFORMATION:
; APPLICANT: BROW, MARY ANN D.
; LYAMICHEV, VICTOR I.
; OLIVE, DAVID M.
; TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
; PATHOGENS
; NUMBER OF SEQUENCES: 165
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/940,925A
; FILING DATE: 10-Jun-2002
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: FORS-01756
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 146:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 620 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 146:
US-09-940-925A-146

```

```

Query Match      11.2%; Score 250.2; DB 11; Length 620;
Best Local Similarity 63.7%; Pred. No. 5.4e-71;
Matches 397; Conservative 0; Mismatches 223; Indels 3; Gaps 1;

QY 916 AACACCTAGCGACCGGACGCGGATACCATCAGGTGGGATAGAGCGGCTGG 975
Db 1 AGCTCGTATGGACCGGACCGGATAGAGCGGATACCATCAGGTGGGATAGAGCGGCTGG 60

QY 976 ACCAAGACCCCTACTCAATGGAGCAATACTTTTGAACCACTCTTTGTTACGAGTGG 1035
Db 61 ACGAACACCCGAGGAATGGGACACAGTTTCTCGAGATCTCTGACGCTACGAGTGG 120

QY 1036 GAGCTTACCAAAAGTCCAGCTGGAGCTTATCAGTGGAAACCAAGACGCGGCGGCT 1095
Db 121 GAGCTGACGAAGAGCGCTGCTGGCGCTTGGCAATACACCGCCAAAGGACGCGCGGTGC 180

QY 1096 GGCACCATACCGGATGCACATGATCCAGCAAGTCGACGCTCCATTTATGCTCATAG 1155

```

```

181 GGCACCATCCGGACCCGTTTCG---GCGGGCAGGGCGCTCCCGACGATCGTCCCACT 237
1156 GACCTGGCGCTGGCGATGACCCCTGATTACGAAAAAATTTCTGACGGTACTATGAAAC 1215
238 GACCTCTCGCTGGCGGTGGATCCGATCTATGAGCGGATCACGCGTCCGCTGGCAAC 297
1216 CTTGATGATGTTGCGATGCTTTTCCGAAAGCATGGTACAAACATGACACAGAGATATG 1275
298 CCGAGGAATTTGGCGGACGAGTTTCGCAAGGCTGGTACAAAGCTGATCCACGAGACATG 357
1276 GGACCAAAAGTTCGCTACTCGGACCAAGTGCCTCAGGAAGACTCTATCTGGCAAGC 1335
358 GGTCCGTTGGAGATACCTTTGGGCGCTGTCCCAAGCAGACCTCTCTGTGGCAGAT 417
1336 CCTATACAGATGTAAGCATCTCTTTGTAGACGAAAAACGATATGAAAGGCTTAAAGCC 1395
418 CCGTCCCTCGGTCAGCACAGACCTCTGTCGCGAAGCCGAGATTGCCAGCTTAAGAGC 477
1396 AATATCCTGGATCGGGACTCACGGTAAGCAGCTGGTGAAGCAGGATGGCTTCTGCA 1455
478 CAGATCCTGGATCGGGATGACTGTCTCAGAGCTAGTTTCGACGCTAGTTTCGACGCTGGCGGCG 537
1456 TCTACTTTTAAAACTCTGACAAGCGCGCGTGCCAAAGCTGCAAGCTATACGACTGCC 1515
538 TCGTGTCTCGGTAGCGACAAGCGCGCGGCCCAACGGTGTGCGCATCCGCTCGCAG 597
1516 CCACAAAAGACTGGGAAGTAAA 1538
598 CCACAAGTCGGTGGGAGTCAA 620

RESULT 10
US-09-940-925A-147/c
; Sequence 147, Application US/09940925A
; Publication No. US20030054338A1
; GENERAL INFORMATION:
; APPLICANT: BROW, MARY ANN D.
; LYAMICHEV, VICTOR I.
; OLIVE, DAVID M.
; TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
; PATHOGENS
; NUMBER OF SEQUENCES: 165
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/940,925A
; FILING DATE: 10-Jun-2002
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: FORS-01756
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 147:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 620 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

```

```

; SEQUENCE DESCRIPTION: SEQ ID NO: 147:
US-09-940-925A-147
Query Match      11.2%; Score 250.2; DB 11; Length 620;
Best Local Similarity 63.7%; Pred. No. 5.4e-71;
Matches 397; Conservative 0; Mismatches 223; Indels 3; Gaps 1;

QY 916 AACACCTACGCGACCGGACACGGTGGGATACCATCACCAGTGGACTAGAGCGGCCCTGG 975
DB 620 AGCTGTATGGCACCGGAACCGGTAAAGGACGCGATCACGACGGCGATCGAGTCTGTATGG 561
QY 976 ACCAAGACCCCTACTCAATGAGCAATAACTTTTTTTGAAACCTCTTTTGGTTACAGTGG 1035
DB 560 ACRAACACCCCGACAAATGGACAACAGTTTCTCGAGATCTCTAGCGGTACAGTGG 501
QY 1036 GAGCTTACCAAAAGTCCAGCTGGAGCTTATCAGTGGAAACCAAAAGAGGGTCCCGGGCT 1095
DB 500 GAGCTGACGAAGAGCCCTGCTGGCGCTTGGCAATACACGCCAAGAGCGGCCGCTGCC 441
QY 1096 GGCACCATACCGGATGACATGATCCACAGCAAGTCCGACGTCCTATTTATGCTCACTAG 1155
DB 440 GGCACCATCCGCGACCCCTTCCG---GCGGGCCAGGGCGCTCCCGACGATGCTGGCCACT 384
QY 1156 GACCTGGCGCTCCGATCGACCCCTGATTTACGAAAAAATTTCTCGACGGTACTATGAAAC 1215
DB 383 GACCTCTCGCTCGGGTGGATCCCGATCTATGAGCGGATCACGCTGCTGGCTGGCAAC 324
QY 1216 CTTGATGATGTTGACAGATGCTTTTCGAAAGCATGGTACAAACTGACACACAGATATG 1275
DB 323 CCGGAGGAATTTGGCGGACGAGTTCCGCAAGGCTGGTACAAAGCTGATCCACGACATG 264
QY 1276 GGACCAAAAGTTCGCTACTCGGACCAAGTGCCTCAGGAAGACTCTATCTATGCGCAAGC 1335
DB 263 GGTCCGTTGGAGATACCTTTGGGCGCTGTGTCGCAAGCAGACCTCTGTGTGGCAGGAT 204
QY 1336 CCTATACAGATGTAAGCATCTCTTTGTAGACGAAAAACGATATTTCAAGGCTTAAAGCC 1395
DB 203 CCGTCCCTCGGTCAGCCACGACCTCTGTCGCGCAAGCGGATTTGCCAGCCTTAAGAGC 144
QY 1396 AAAATCCTGGAATCGGACTGACGGTAAAGCAGCGTGGTAAAGCAGCGCATGGCTTCTGCA 1455
DB 143 CAGATCCGGCATCGGATTTGACTCTCTACAGCTAGTTTCACCGCATGGCGGGCGG 84
QY 1456 TCTACTTTTAAACTCTGACAAGCGCGCGTCCGCAAGGTCGCAAGGTCAGATATACGACTGGCC 1515
DB 83 TCGTGTCTCCGTGGTAGCGACAAGCGCGCGCGCAAGCGTGGTGGCATCCGCTCGCAG 24
QY 1516 CCACAAAAGACTGGGAAGTAAA 1538
DB 23 CCACAAGTCGGTGGGAGTCAA 1

RESULT 11
US-09-940-925A-150/c
; Sequence 150, Application US/09940925A
; Publication No. US20030054338A1
; GENERAL INFORMATION:
; APPLICANT: BROW, MARY ANN D.
; LYAMICHEV, VICTOR I.
; OLIVE, DAVID M.
; TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
; PATHOGENS
; NUMBER OF SEQUENCES: 165
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/09/940,925A
:   FILING DATE: 10-Jun-2002
:   CLASSIFICATION: <Unknown>
: ATTORNEY/AGENT INFORMATION:
:   NAME: CARROLL, PETER G.
:   REGISTRATION NUMBER: 32,837
: REFERENCE/DOCKET NUMBER: FORS-01756
: TELECOMMUNICATION INFORMATION:
:   TELEPHONE: (415) 705-8410
:   TELEFAX: (415) 397-8338
: INFORMATION FOR SEQ ID NO: 150:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 620 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: double
:   TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: SEQUENCE DESCRIPTION: SEQ ID NO: 150:
US-09-940-925A-150

```

```

Query Match      11.2%; Score 250.2; DB 11; Length 620;
Best Local Similarity 63.7%; Pred. No. 5.4e-71;
Matches 397; Conservative 0; Mismatches 223; Indels 3; Gaps 1;

QY 916 AACACCTACGGCACCAGGACACAGGTGGGATACCATCACCAGTGGAGCTAGAAAGCGCGCTGG 975
Db 620 AGCTCGTATGGCACCAGGACCGTGAAGAGCGGATCACCACCGGATCGAGGTGGTATGG 561
QY 976 ACCAAGACCCCTACTCAATGAGCAATACTTTTGAACACCTCTTTGTTACGAGTGG 1035
Db 560 ACGAACACCCCGAGCAAGTGGCAACAGTTTCTCGAGATCTGTACGGCTACGAGTGG 501
QY 1036 GAGCTTACCAAAAGTCCAGTGGAGCTTATCAGTGGAAACCAAAAGACGGTGGCGGGCT 1095
Db 500 GAGCTGACGAAGAGCCCTGCTGGCGTTGCCATACACCCCAAGACGGCGCGGTGGC 441
QY 1096 GGCACATACCGGATGACATGATCCAGCAAGTCCGAGCTCCATTTATGCTACTACG 1155
Db 440 GGCACCATCCGAGCCGCTTCG---GCGGGCCAGGGCGCTCCCGACGATGCTGGCCACT 384
QY 1156 GAGCTGGCGTGGCGATGACCTGATACGAAATAATTTCTCGAGGTTACTATGAAC 1215
Db 383 GAGCTCTCGTGGGTGGATCGGATCTATGACGGGATCAGCGGTGCTGGCTGGGAACAC 324
QY 1216 CTGTATGAGTTTGCAGATGCTTTCGGGAAAGCATGTTCAAACTGACACACAGATATG 1275
Db 323 CCGGAGGAATTGGCCGACGAGTTTCGCAAGGCCCTGTACAAGCTGATCCACCGAGATG 264
QY 1276 GGACCAAAAGTGGCTTACCTGGGACCAAGAGTCCCTCAGAAAGACCTCATCTGGCAAGC 1335
Db 263 GGTCCGTTGCGAGATACCTTGGGCGCTGGTCCCAAGCAGACCCCTGTGTGGCAGGAT 204
QY 1336 CTATACAGATGTAAGCCATCTCTGTAGAGCAAAACGATATTCAGGSCCTTAAGGC 1395
Db 203 CCGGTCCCTTGGGTGACGACGACCTCTGTCGGGGAAGCGGAGATCCGAGCCCTTAAGAGC 144
QY 1396 AAAATCCTGGAATCGGAGTACGGTAAAGAGCTGGTAAAGCAGCGCATGGGCTTCTGCA 1455
Db 143 CAGATCCTGGCATCGGATTGACTGTCTACAGCTAGTTTCGACCCGATGGGCGGCGG 84
QY 1456 TCTACTTTTAGAACTCTCACAAGCGCGGGTGGCCAGCGGTGACGATGACGATGGCC 1515
Db 83 TGTGCTTCCGTTGAGCAGCAAGCGCGGCGGCAAGCGGTGGTGGCATCGGCTGCGAG 24
QY 1516 CCACAAAAGACTGGGAAGTAAA 1538
Db 23 CCACAAGTCCGGTGGGAGGTCAA 1

```

RESULT 12

```

US-09-941-193A-143
: Sequence 143, Application US/09941193A
: Publication No. US2003010873A1
: GENERAL INFORMATION:
: APPLICANT: BROW, MARY ANN D.
: LYAMICHEV, VICTOR I.
: OLIVE, DAVID M.
: TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
: PATHOGENS
: NUMBER OF SEQUENCES: 165
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MEDLEN & CARROLL
: STREET: 220 MONTGOMERY STREET, SUITE 2200
: CITY: SAN FRANCISCO
: STATE: CALIFORNIA
: COUNTRY: UNITED STATES OF AMERICA
: ZIP: 94104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/941,193A
: FILING DATE: 28-Aug-2001
: CLASSIFICATION: <Unknown>
: ATTORNEY/AGENT INFORMATION:
:   NAME: CARROLL, PETER G.
: REGISTRATION NUMBER: 32,837
: REFERENCE/DOCKET NUMBER: FORS-01756
: TELECOMMUNICATION INFORMATION:
:   TELEPHONE: (415) 705-8410
:   TELEFAX: (415) 397-8338
: INFORMATION FOR SEQ ID NO: 143:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 620 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: double
:   TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: SEQUENCE DESCRIPTION: SEQ ID NO: 143:
US-09-941-193A-143

```

```

Query Match      11.2%; Score 250.2; DB 11; Length 620;
Best Local Similarity 63.7%; Pred. No. 5.4e-71;
Matches 397; Conservative 0; Mismatches 223; Indels 3; Gaps 1;

QY 916 AACACCTACGGCACCAGGACACAGGTGGGATACCATCACCAGTGGAGCTAGAAAGCGCGCTGG 975
Db 1 AGTCTGTATGGCACCAGGACCGGTAAGAGCGGATCACCAGCGGATCGAGGTGCTATGG 60
QY 976 ACCAAGACCCCTACTCAATGAGCAATACTTTTGAACACCTCTTTGTTACGAGTGG 1035
Db 61 ACGAACACCCCGAGCAAGTGGGACACAGTTTCTCGAGATCTGTACGGCTAGAGTGG 120
QY 1036 GAGCTTACCAAAAGTCCAGTGGAGCTTATCAGTGGAAACCAAAAGACGGTGGCGGGCT 1095
Db 121 GAGCTACGAAAGAGCCCTGCTGGCGTTGGCAATACACGCCCAAGGACGCGCGCTGGC 180
QY 1096 GGCACCATACCGGATGACATGATCCAGCAAGTCCGACGCTCCATTTATGCTCACTAGC 1155
Db 181 GGCACCATCCGAGNCCCGTTGCG---GCGGGCCAGGGCGCTCCCGACGATGCTGGCCACT 237
QY 1156 GAGCTGGCGTGGCGATGGACCTGATACGAAAAAATTTCTCGACGGTACTATGAAAC 1215
Db 238 GAGCTCTCGTGGGTGGATCCGATCTATGAGCGGATCAGCGGTGCTGGCTGGGACAC 297
QY 1216 CTGTATGAGTTTGCAGATGCTTTCGGGAAAGCATGTTCAAACTGACACACAGATATG 1275
Db 298 CCGGAGGAATTGGCCGACGAGTTTCGCAAGGGCTGGTACAAGCTGATCCACCGAGACATG 357
QY 1276 GGACCAAAAGTGGCTTACCTGGGACGAGTGGCTCAGGAAGACCTCATCTGGCGAAGC 1335

```

358	GGTCCCGTTCGGAGATACCTTGGCCCGCTGGTGGCCCAAGACAGACCTGCTGTGGCAGGAT	417
1336	CCTATACCAAGATGAAGCCATCCTCTGTAGACGAAACGATATTTGAAGGCCTTAAAGGC	1395
418	CGGTCCCTCGGTACAGCCAGCAGCTCGTCGGCGAAGCCGAGATGCCAGCCTTAAAGGC	477
1396	AAATCCTGGAATCGGGACTGACGGTAAGCGAGCTGGTAAGACGCGATGGGCTTCTGCA	1455
478	CAGATCGGGCATCGGGATGACTGTCTCACAGCTAGTTTCGACCGCATGGCGCGCGG	537
1456	TCTACTTTTAGAAACTCTGCACAAAGCGCGGGTGCACCGGTGCAGCTATACGACTGGCC	1515
538	TCGTCGTTCCGTGGTAGCAGACGCGGGCGGCCCAACGGTGTGCATCCGCTGCAG	597
1516	CCACAAAAGACTGGGAAGTAA	1538
598	CCACAAAGTCGGTGGGAGGTCAA	620

RESULT 13

US-09-941-193A-146

; Sequence 146, Application US/09941193A

; Publication No. US20030108873A1

; GENERAL INFORMATION:

; APPLICANT: BROW, MARY ANN D.

; LYAMICHEV, VICTOR I.

; OLIVE, DAVID M.

; TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF PATHOGENS

; NUMBER OF SEQUENCES: 165

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MEDLEN & CARROLL

; STREET: 220 MONTGOMERY STREET, SUITE 2200

; CITY: SAN FRANCISCO

; STATE: CALIFORNIA

; COUNTRY: UNITED STATES OF AMERICA

; ZIP: 94104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/941,193A

; FILING DATE: 28-Aug-2001

; CLASSIFICATION: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: CARROLL, PETER G.

; REGISTRATION NUMBER: 32,837

; REFERENCE/DOCKET NUMBER: FORS-01756

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 705-8410

; TELEFAX: (415) 397-8338

; INFORMATION FOR SEQ ID NO: 146:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 620 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; SEQUENCE DESCRIPTION: SEQ ID NO: 146:

US-09-941-193A-146

Query Match	11.2%	Score 250.2;	DB 11;	Length 620;
Best Local Similarity	63.7%	Pred. No. 5.4e-71;		
Matches 397; Conservative	0;	Mismatches 223;	Indels 3;	Gaps 1;
Qy 916 AACACCTACGGCACCGGACACGGTGGCGATACCATCACAGTGACTAGAACGGCGCTGG	975			
Db 1 AGTCGTATGGCACCGGAACCGGTAAGGACGGCATCACCGGCGATCGAGGTCGTATGG	60			
Qy 976 ACCAAGACCCCTACTCAATCGACGACATACTTTTTTGAACAACTCTTTGGTTACGAGTGG	1035			

Db	61	ACGAACACCCCGACGAAATGGGAAACAGATTCTCTCGAGATCCTGTACGGCTACGAGTGG	120
Qy	1036	GAGCTTACCAAAAGTCCAGCTGAGCTTATCAGTGGAAACCAAAAGACGGTCCGGGGCT	1095
Db	121	GAGCTGACGAAGAGCCCTGCTGGCGCTTGGCATACACCGCCAAAGGACGGCGCGGTGCC	180
Qy	1096	GGCACCATCCGGATGCACATGATCCAGCAAGTCGACAGCTCCATTTATGTCTACTAGG	1155
Db	181	GGCACCATCCGGACCGCTTCG---CGCGGCCAGGGCGCTCCCGGACGATGCTGGCCACT	237
Qy	1156	GACCTGGCGCTCGCATGGACCTGATTACGAAAAAATTTCTCGACGGTACTATGAAC	1215
Db	238	GACCTTCGCTGGGGTGGATCCGATCTATGACGGATCAGCGCTCGCTGGGTGGAACAC	297
Qy	1216	CTGTGATGATTTGCAGATGCTTTTCGGGAAAGCATGTACAAACTGACACACAGACATATG	1275
Db	298	CCCGAGGAAATTGCCGAGAGTTTCGCAAGGCCCTGCTACAAGCTGATCCACCGAGACATG	357
Qy	1276	GGACCAAGTGGCTTACCTTGGGACCAAGATGCCCTCAGGAAGACCTCATCTGTGGCAAGAC	1335
Db	358	GGTCCCGTTTCGAGATACCTTGGCGCGCTGGTCCCAAGCAGACCCCTGCTGTGGCAGGAT	417
Qy	1336	CCATATACCATGTAAGCCATCTCTGTGTAGCAAAACGATATTGAAGGCCCTAAAGCC	1395
Db	418	CCGGTCCCTCGGTTCAGCCAGACCTCGTCGGCGAAGCCGAGATGCCAGCCCTTAAGAC	477
Qy	1396	AAATCTGGAATCGGGACTGACGGTAAGCGAGCTGTAAACACGCGATGGGCTTCTGCA	1455
Db	478	CAGATCTGCAATCGGGATTTGACTGTCTACACCTAGTTTCGACCCGATGGCGGGCGG	537
Qy	1456	TCCTACTTTTGAACCTCTGCAAGCGCGGGTGGCCAAAGGTGCGACGTATAGACCTGGCC	1515
Db	538	TCGTCGCTCGTGTAGCAGAAAGCGCGGGCGGCGCAACGGTGGTGGCATCCGCTTGCAG	597
Qy	1516	CCACAAAAGACTGGGAAGTAAA	1538
Db	598	CCACAGCTGGGTGGGAGGTCAA	620

RESULT 14

US-09-941-193A-147/c

; Sequence 147, Application US/09941193A

; Publication No. US20030108873A1

; GENERAL INFORMATION:

; APPLICANT: BROW, MARY ANN D.

; LYAMICHEV, VICTOR I.

; OLIVE, DAVID M.

; TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF

; PATHOGENS

; NUMBER OF SEQUENCES: 165

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MEDLEN & CARROLL

; STREET: 220 MONTGOMERY STREET, SUITE 2200

; CITY: SAN FRANCISCO

; STATE: CALIFORNIA

; COUNTRY: UNITED STATES OF AMERICA

; ZIP: 94104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/941,193A

; FILING DATE: 28-Aug-2001

; CLASSIFICATION: <unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: CARROLL, PETER G.

; REGISTRATION NUMBER: 32,837

; REFERENCE/DOCKET NUMBER: FORS-01756

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 705-8410

; TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 147:
SEQUENCE CHARACTERISTICS:
LENGTH: 620 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 147:
US-09-941-193A-147

Query Match 11.2%; Score 250.2; DB 11; Length 620;
Best Local Similarity 63.7%; Pred. No. 5.4e-71;
Matches 397; Conservative 0; Mismatches 223; Indels 3; Gaps 1;

QY 916 AACACCTACGCGACCGGACGCGGATACCATACCATGACGAGGAGGCGGCTGG 975
Db 620 AGCTCGTATGCGACCGGAGCGTAAGGACGCGATACCATGCGGATGCTGGTATGG 561
QY 976 ACCAAGACCCCTACTCAATGAGCAATACTTTTGAACACCTCTTTGGTTACGATGG 1035
Db 560 ACGAACCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 501
QY 1036 GAGCTTACCAAAAGTCCAGTGGAGCTTATCAGTGGAAACCAAAAGAGCGTCCGGGGCT 1095
Db 500 GAGCTGACGAGAGCCCTGCTGGGCGCTTGGCAATACACCGGAGGAGGAGGAGGAGG 441
QY 1096 GGCACCATACCGGATGACATGATCCAGCAAGTCCGACGCTCCATTTATGCTCAGTACG 1155
Db 440 GGCACCATACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 384
QY 1156 GACCTGGCGCTGCGCATGGACCCCTGATTACGAAAAAATTTCTGACGCTACTATGAAAC 1215
Db 383 GACCTCTCGCTGCGGATGATCCGATCTATGAGGCGGATCAGGCTCGCTGGCTGGAAAC 324
QY 1216 CCTGATGAGTTGCGAGATGCTTTGCGAAAGCATGTTACAACTGACACACAGATATG 1275
Db 323 CCGGAGGAAATGGCCGACGAGTTGCGCAAGGCGCTGTTACAAGCTGATCCACCGAGATG 264
QY 1276 GGACCAAGGTGGCTACCTGGGACCGGAGTCCGAGCAAGTCCGAGGAGGAGGAGGAGG 1335
Db 263 GGTCCCGTTCGAGATACCTTGGGCGGCTGGTCCCAAGCAGAGACCTGCTGTGGCAGAT 204
QY 1336 CCTATACGAGTGAAGCCTCTCTTTGAGCAAAACGATATGAAAGGCTTAAAGGC 1395
Db 203 CCGCTCTCGCTGCGGATGATCCGATCTATGAGGCGGATCAGGCTCGCTGGCTGGAAAC 144
QY 1396 AAATCCTGGAATCGGAGTGGAGCTGTTGAGCAAGTGGTAAAGCAGGAGGAGGAGGAGG 1455
Db 143 CAGATCGGAGATGGCCGACGAGTTGCGCAAGGCGCTGGTACAAGCTGATCCACCGAGATG 84
QY 1456 TCTACTTTTGAACCTCTGACAGCGGCGGCTGGCCAAAGGAGGAGGAGGAGGAGGAGG 1515
Db 83 TCGTCTCGCTGGTGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 24
QY 1516 CCACAAAGACATGGGAAGTAAA 1538
Db 23 CCACAGTCCGGTGGGAGGTCAA 1

RESULT 15

US-09-941-193A-150/c
Sequence 150, Application US/09941193A
Publication No. US20030108873A1

GENERAL INFORMATION:

APPLICANT: BROW, MARY ANN D.
LYAMICHEV, VICTOR I.
OLIVE, DAVID M.

TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF

NUMBER OF SEQUENCES: 165

CORRESPONDENCE ADDRESS:

ADDRESS: MEDLEN & CARROLL

STREET: 220 MONTGOMERY STREET, SUITE 2200

CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/941,193A

FILING DATE: 28-Aug-2001

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: CARROLL, PETER G.

REGISTRATION NUMBER: 32,837

REFERENCE/DOCKET NUMBER: FORS-01756

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 150:

SEQUENCE CHARACTERISTICS:

LENGTH: 620 base pairs

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 150:

US-09-941-193A-150

Qy 1516 CCACAAAGAGCTGGGAAGTAA 1538
||||| | ||||| ||
Db 23 CCACAAGTCGGGTGGGGAGTCAA 1

Search completed: October 8, 2003, 19:55:28
Job time : 577.431 secs

10/10/2003 11:39:17 P. 13/13

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2003, 07:49:13 ; Search time 4638.63 Seconds
(without alignments)
11726.187 Million cell updates/sec

Title: US-09-884-889-7
Perfect score: 2238
Sequence: 1 atggaatacacaaacacac.....accggtttgatctgaaataa 2238

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mus:*
23: em_gss_mam:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	296.2	13.2	786	28	BH403595
2	283	12.6	1102	28	AF075865
3	269	12.0	873	14	CD377714
4	261	11.7	565	28	BH401331

5	259.8	11.6	729	9	AW179968
6	256.4	11.5	762	13	BQ752115
7	254.8	11.4	707	13	BQ751801
8	249.4	11.1	710	13	BU065388
9	244.6	10.9	918	28	AF075904
10	243	10.9	737	29	CNS01HHH
11	242.8	10.8	742	13	BQ751881
12	239	10.7	562	10	BG370112
13	237.6	10.6	663	12	BM865076
14	233.2	10.4	639	13	BQ751512
15	231.2	10.3	633	13	BQ752114
16	229.2	10.2	606	13	BQ751387
17	225.8	10.1	757	14	CD425096
18	223	10.0	573	9	AW180683
19	216.6	9.7	628	28	AQ000148
20	216.2	9.7	701	14	CD422961
21	215.4	9.6	745	14	CD425011
22	213.8	9.6	757	14	CD427691
23	213.4	9.5	666	28	AQ324744
24	212	9.5	447	28	BH387681
25	208.4	9.3	656	13	BU062488
26	208.4	9.3	741	14	CD426471
27	205.8	9.2	712	13	BU062489
28	204.6	9.1	718	12	BI750328
29	202.4	9.0	646	28	AQ399314
30	200.8	9.0	660	13	BQ751388
31	199.6	8.9	571	9	AW180185
32	198	8.8	945	14	CD459390
33	195.8	8.7	465	10	BG278355
34	194.8	8.7	600	9	AW180270
35	189.4	8.5	598	14	CD426092
36	186.2	8.3	613	13	BQ751511
37	184.8	8.3	531	13	BQ143330
38	182.2	8.1	546	13	BQ142683
39	181.2	8.1	472	9	AA901970
40	177.4	7.9	607	14	CD461624
41	175	7.8	495	10	BE759355
42	174.2	7.8	627	9	AW180016
43	174	7.8	700	9	AW180906
44	170	7.6	601	10	BE758793
45	169	7.6	588	14	CD457487

ALIGNMENTS

RESULT 1	BH403595	786 bp	DNA	linear	GSS 11-DEC-2001				
LOCUS	AG-ND-127D17.TF	ND-TAM	Anopheles gambiae	genomic clone	AG-ND-127D17				
DEFINITION	, genomic survey sequence.								
ACCESSION	BH403595								
VERSION	BH403595.1	GI:17349811							
KEYWORDS	GSS.								
SOURCE	Anopheles gambiae	(African malaria mosquito)							
ORGANISM	Anopheles gambiae								
REFERENCE	1 (bases 1 to 786)								
AUTHORS	Hong, Y.S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J., Ren, C., Huff, E.R., Carlile, J.L., Black, K., Zhang, H.-B., Gardner, M.J. and Collins, F.H.								
TITLE	Construction of a BAC library and generation of BAC end sequence-tagged connectors for genome sequencing of the African malaria mosquito Anopheles gambiae								
JOURNAL	Mol. Genet. Genomics 268 (6), 720-728 (2003)								
MEDLINE	22542063								
PUBMED	12655398								
COMMENT	Other_GSSs: AG-ND-127D17.TR Contact: Brendan J Loftus Department of Eukaryotic Genomics The Institute for Genomic Research								

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: bjoftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.

Seq primer: M13 For

Class: BAC ends.

FEATURES

Location/Qualifiers

1..786

/organism="Anopheles gambiae"

/mol_type="genomic DNA"

/strain="PEST"

/db_xref="taxon:7165"

/clone="AG-ND-127D17"

/clone_lib="ND-TAM"

/note="Vector: pECBAC1; Site_1: HindIII"

183 a 211 c 148 g 244 t

BASE COUNT

ORIGIN

Query Match 13.2%; Score 296.2; DB 28; Length 786;

Best Local Similarity 66.5%; Pred. No. 1.2e-75;

Mismatches 490; Conservative 0; Mismatches 233; Indels 14; Gaps 4;

```

QY 505 CTCACAGGAAGCTAGCTCTGGAACATATGGCTTTAAACATTTGGT-TTTCAGAGTGG 563
DB 728 CTAACCGGTAAAGTTGCTTTGGAATCTATGGATACAAACCAAGGATTTGGCGGTGC 669
QY 564 CAGACGAGATGATGGAGCCCTGAAGAAGATGTATCTGGGA- - -GCAGAAACCAAT 619
DB 668 CCGTACAGATGTATGGAAACCGGATTAAGATGTATCTTGGGAAGATGAGAAACCTGGC 609
QY 620 GGCTGGGAGACAGCGCTATGAAGTACCGAGAGCTCGAAATCCCTGGGAGCCGTAC 679
DB 608 TGGGAACACATAAAGATATACAGGGATCGTGATCTGCAAGATCGCTGGCAGCAACA 549
QY 680 AATATGGAGTCTATGTAAACCGGAGGAGCCCAACGGCAAGCCAGACCCCTATCGCTG 739
DB 548 CAATGGGCTTATCTATGTAAATCTGAAGGACCGGAGGAAACCGGATCGCTGGCAG 489
QY 740 CTGCGCGTGATATCTGAGACTTTTGGCCGAATGGAATGAATGACGAAGAACCGTGG 799
DB 488 CTGCTAAAGATATCGGTGATACATTTGGCCGTATGGGAATGATGACGAGGAACGTAG 429
QY 800 CTCTCATAGCGGTGGACACACTTGGGAAACCCCATGGTCTGCCGATCGGAGAAAT 859
DB 428 CGCTAATTCGTGTGGCCACACATTTGGTAAACATCATGGTCCGAGATGACAGCGCTA- 370
QY 860 ATGTGGGCGGACAGCTCGCCCGCAGGTATTGAAGAATGAGCTGGGTTGGGAGAAACA 919
DB 369 -GTAGGAGCTGAGCGCTGAAGATCCGGAATTTGAAGCTCAGGGATTAGTGGGCAAGTA 312
QY 920 CTACGGCACCGGACAGCGTGGGATACCATCAGCAGTGGACTAGAGAAGCGCTGGACCA 979
DB 311 AATTTGGAAACAGCTCCGAGCAGATGCTATTACCATGTTGCTGGAAGTTTACCTGGACCA 252
QY 980 AGACCCCTACTCAATGGAGCAATAACTTTTTTGAACACCTCTTTGGTACGAGTGGGAGC 1039
DB 251 AAACACCAAGAAATGGACCAAGCACTTCTTCAGCAGCTTATTCAGCTATGATGGAGC 192
QY 1040 TTACCAAAAGTCCAGCTGAGCTTATCAGTGGAAACCAAAAGACGGTGGCCGGGTGGCA 1099
DB 191 TAACGAAAGACCGCTGAGCTGCTCATCAGTGGATTGCAAAAGA- - -TGCCGGAGATA 138
QY 1100 CCATACCGGATGCATGATGCCAGCAAGTCGACAGCTCCATTTATGCTACTACGGACC 1159

```

```

Db 137 TTAATCCGAATGCTTTTGTATGCAGAGAAAAAGGTCGCCGTATATGCTTACAACTGACC 78
QY 1160 TGGCGCTGGGCATGGACCTCTGATTACGAAAAAATTTCTCGACGCTACTATGAAAAACCTG 1219
Db 77 TTTCAATTAAGTTTGGTCTTATTTATGAGAAATATCCAGACGCTTCTATGAAAAATCCGG 18
QY 1220 ATGAGTTTGCAGATGCT 1236
Db 17 AAGAACTGGCAGAGACT 1

RESULT 2
AF075865 1102 bp DNA linear GSS 29-AUG-2000
LOCUS AF075865 Salmonella typhimurium LT2, Lambda DASH II Salmonella
DEFINITION typhimurium genomic clone 1079-T7, genomic survey sequence.
ACCESSION AF075865
VERSION AF075865.1 GI:3320735
KEYWORDS GSS.
SOURCE Salmonella typhimurium
ORGANISM Salmonella typhimurium
REFERENCE 1 (bases 1 to 1102)
AUTHORS Wong,R.M.-Y., Wong,K.K., Benson,N.R. and McClelland,M.
TITLE Sample sequencing of a Salmonella typhimurium LT2 lambda library:
JOURNAL comparison to the Escherichia coli K12 genome
MEDLINE FEMS Microbiol. Lett. 173 (2), 411-423 (1999)
PUBMED 99243757
COMMENT 10227170
Contact: McClelland M
Molecular Biology
Sidney Kimmel Cancer Center
3099 Science Park Road, San Diego, CA 92121, USA
Email: mclelland@lifesci.sdsu.edu
Class: shotgun.
Location/Qualifiers
1..1102
/organism="Salmonella typhimurium"
/mol_type="genomic DNA"
/strain="LT2"
/db_xref="taxon:602"
/clone="1079-T7"
/clone_lib="Salmonella typhimurium LT2, Lambda DASH II"
/note="Vector: Lambda DASH II; sequenced using Li-Cor
sequencer"
BASE COUNT 235 a 278 c 334 g 251 t 4 others
ORIGIN

Query Match 12.6%; Score 283; DB 28; Length 1102;
Best Local Similarity 58.6%; Pred. No. 1.2e-71;
Matches 595; Conservative 0; Mismatches 389; Indels 31; Gaps 5;

QY 1225 TTTGCAGATGCTTTCGGAAGCATGGTACAACTCACACAGAGATATGGACCAAG 1284
DB 17 TTTAATGAAGCTTTGCTCGTGGTTCAAACTGACCAAGATATGGACCAAAA 76
QY 1285 GTGCGCTACTGGGACCAAGTGCCTCAGGAAGACCTCATCTGGAAGACCTATACCA 1344
DB 77 GCGCGTTACATCGGACCGGAAGTGCAGGAAGAGATCTGATCTGGCAGGACCGCTTGGCG 136
QY 1345 GATGTAAGCCATCTCTTCTTAGACAAAACGATATTGAAGGCTTAAAGCCAAAATCTTG 1404
DB 137 CAACCGCTCTATCAGCAACGCGAGGAAGACATATCAAA- - -CCTGAAGCGCGATCGCT 193
QY 1405 GAATCGGACTGACGGTAAAGCGAGCTGGTAAGCAGCGGATGGCTTCTGCATCTACTTTT 1464
DB 194 GCATCGGGCTTTCTATTAGCAGATGGTTTCGGTTTCGGCTGGGCATCCCGCTACTTTC 253
QY 1465 AGAAACTCTGACAAAGCGCGGTGCAACGGTGGACGTATACGACTGCCCCCACAAAAA 1524
DB 254 CCGGCGGCGGATTAAGCTTGGCGCGCTAACGGCGCGCTGTGGCATTTAGCGCTACGCC 313

```



```

|||||
781 TTGAAAAA 789

DB
RESULT 4
BH401331
LOCUS
DEFINITION AG-ND-137N19.TR.1 ND-TAM Anopheles gambiae genomic clone
AG-ND-137N19, genomic survey sequence.
ACCESSION BH401331
VERSION BH401331
KEYWORDS GSS,
SOURCE BH401331.1 GI:17347547
ORGANISM Anopheles gambiae (African malaria mosquito)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoides;
Anopheles.
REFERENCE 1 (bases 1 to 565)
AUTHORS Hong, Y.S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J., Ren
, C., Huff, E.R., Carlile, J.L., Black, K., Zhang, H.-B., Gardner, M.J.
and Collins, F.H.
TITLE Construction of a BAC library and generation of BAC end
sequence-tagged connectors for genome sequencing of the African
malaria mosquito Anopheles gambiae
Mol. Genet. Genomics 268 (6), 720-728 (2003)
JOURNAL 22542063
MEDLINE 12655398
PUBMED
COMMENT Other_GSSs: AG-ND-137N19.TF.1
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjl@fugate.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seq primer: M13 Rev
Class: BAC ends.
Location/Qualifiers
1..565
/organism="Anopheles gambiae"
/mol_type="genomic DNA"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-137N19"
/clone_lib="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT 173 a 104 c 145 g 143 t
ORIGIN

Query Match 11.7%; Score 261; DB 28; Length 565;
Best Local Similarity 66.4%; Pred. No. 2.5e-65;
Matches 375; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

QY 1238 TCCGGAAGCATGGTACAACTGACACACAGAGATATGGGACCAAGGTCCGTACTCGG 1297
DB 1 TTGCTAAGCATGGTTCAAACTGTTCCACCGCTATATGGGGCCCAAAATCAAGATCCTG 60
QY 1298 GACCAGAGTGCCTCAGGAGACCTCATCTGGCAAGACCTATACCATGATGAAGCATC 1357
DB 61 GACCAGAGTTCGGCGAAGACCTTATCTGGCAGGATCCTATTCGGCGAGTAGATCATC 120
QY 1358 CTCTTGTAGCAAAACGATATTGAAGGCTCAAAAGCCAAAATCCCTGGAATCCGGACTGA 1417
DB 121 AGCTGATCGATGCTCAGGATGAGGCACAGCTAAAAAATAAGTATTGGATTCCGGATTAT 180

```

```

QY 1418 CGGTAAGCGAGCTGGTAAAGCAGCGATGGGCTTCTGCATCTACTTTTAAAGAACTCTGACA 1477
DB 181 CAGTTTCTGAACTGGTATCTACAGCATGGGCTTCTGCTTCTACATTCGAGGATCGGATA 240
QY 1478 AGCCGGGGGTGCCAAGCGGTGCAGTATACGACTGGGCCCCACAAAAGAGACTGGGAAGTAA 1537
DB 241 AAAGAGGAGGTGCAAAATGGTGCAGTATTCGTTTGGAGCCACAAAGAAGCTGGGCGGTAA 300
QY 1538 ACAACCCCTCAGCAACTTCCAGGGTACTCAAAACACTAGAGAGTATCCAGGAGGACTTTA 1597
DB 301 ACAATCCGGAACAGTTGTCAAAGTATTGAATGATTGGGAAGGTATTTCAGAAAGAGTTCA 360
QY 1598 ACCAGGGCCATCAGATAAACAAGCAGTATCGTTGGCGCAGCTGATGTGTGCTGCGCGGCT 1657
DB 361 ATGCAGCAGCAAACTGGAGGCCAAAAGTTTCATTATGACAGACCTTATTGTCTTGCAGGAA 420
QY 1658 GTGCGGGGTGTAGAAAAGCTGCAAAAGAGTGTGGCCATGAGGTGCAGGTGCTTTTCAACC 1717
DB 421 ATGCAGGTGTAGAAAAGCAGCAGCAAAATGACAGGTGTGACGGTAAATGTTCCGTTACTG 480
QY 1718 CGGCAGCAGCGGATGCCACCGCTGACAAACCGATGTGGAGCTTTTCGAAGCAGCTAGAC 1777
DB 481 CAGACGTATGGATGTCATCACAAGCAGCAGTGTATGATAGATCTTCTCTTATTATTAGAGC 540
QY 1778 CAGCGGCTGACGGCTTTAGAACTA 1802
DB 541 CTTATGCAGATGGTTCCGTAACTA 565

RESULT 5
AW179968
LOCUS
DEFINITION MGA0020f Mga Library Mycosphaerella graminicola linear EST 17-NOV-1999
5', similar to catalase-peroxidase, mRNA sequence.
ACCESSION AW179968
VERSION AW179968.1 GI:6447163
KEYWORDS EST.
SOURCE Mycosphaerella graminicola
ORGANISM Mycosphaerella graminicola
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes et
Chaetothryomycetes incertae sedis; Mycosphaerellaceae;
Mycosphaerella.
REFERENCE 1 (bases 1 to 729)
AUTHORS Keon, J.P.R., Bailey, A.M. and Hargreaves, J.A.
TITLE A group of expressed cDNA sequences from the wheat fungal leaf
blotch pathogen, Mycosphaerella graminicola (Septoria tritici)
JOURNAL Fungal Genet. Biol. 29 (2), 118-133 (2000)
MEDLINE 20374020
PUBMED 10919380
COMMENT Contact: Hargreaves JA
Cell Biology Department
IACR-Long Ashton Research Station
Long Ashton, Bristol, BS41 9AF, UK
Tel: +44(0)1275 392181
Fax: +44(0)1275 394281
Email: John.hargreaves@bbsrc.ac.uk
Insert Length: 2600 Std Error: 0.00
Seq primer: M13 reverse.
Location/Qualifiers
1..729
/organism="Mycosphaerella graminicola"
/mol_type="mRNA"
/strain="Strit"
/db_xref="taxon:54734"
/clone="Mga0020"
/clone_lib="Mga Library"
/note="Vector: pSPORT1; Library constructed from cultures
utilizing ammonium ions as a source of nitrogen"
BASE COUNT 159 a 217 c 203 g 150 t
ORIGIN

Query Match 11.6%; Score 259.8; DB 9; Length 729;

```

```

Best Local Similarity 64.6%; Pred. No. 6.6e-65;
Matches 387; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

QY 90 TCAGGTGGCGCACCACAAAGAGGATGGTGGCCCAACATGCTCAACCTCGGCATCTT 149
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 82 TGGCGGTGGTGTACCCGCAACAGGAGCTGGTGGCCAAATGAGCTGAACCAACAGATCCT 141
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 150 AGCCCAACATTCATCGCTATCGGACCCCAACACCGGATTTTGTACTATGCCGAAGATT 209
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 142 GGGCCAGCAGCAGCTGTGTACCGCCATTCGCGAAGCAGTTCGACTCCACCGCATTT 201
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 210 TAAGAAGCTAGATCTGGCAGCGGTTAAAAAGACCTGGCAGCGCTTAATACACAGATTCA 269
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 202 CAAGAGCTTGGATTACATGGCTCAAGAGAGCTTTGAACGACCTCATGACCGATCGAA 261
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 270 GGACTGGTGGCGCAGAGATACGGCTATTATGGCCCTTTCTTTATACGATGGCGTGGCA 329
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 262 GGACTTCTGGCCAGCAGACTTTGGCCACTTACGGTGGTTTCTTCTCGCATGGCATGGCA 321
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 330 CAGCGCGCAGCTACCGTATCGGTATGGCGCGTGGTGGCGTGGCTCGGCTCACAGCG 389
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 322 CAGCGCTGTGATACCGCTGCTGATGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 381
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 390 CTTGCGGCTCTCAATAGCTGGCCAGACAAATGCCAATCTGGATAAGCAGCTTGTCTCT 449
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 382 ATTGCGTCTCTCAACTATGGCGGACAAAGCCCAATCTCGACAGGCTCGCGCGCTGT 441
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 450 TTGGCCCAATCAACAAATAACGGTGGAAATATCTCTGGGGGATCTTAATGATCTAC 509
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 442 GTGGCAATCAAGCAAAAGTACGAAACAAAGATCTCATGGCGGACCTGTCTCTCTAC 501
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 510 AGGAACGTAGCTCTGGAACATATGGGCTTTAAACTTTTGGTTTGGAGTGGGAGAGC 569
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 502 TGAATGTGCGCAATGATCTATGGTCTGCCACGTTTGGTTTGGCGGGTGGCGTGC 561
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 570 AGATGTATGGGAGCTTGAAGAAGATCTATCTAGTGGGAGCAGAAACGAATGCTGGAGA 629
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 562 GCACATATGGGAGGAGTATCTCGTATCTAGTGGGTTGGTGGAGCACCCTTTTGGGCA 621
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 630 CAAGCGCTATGAAGTACCGAGAGTCCGAAATCCCTGGGAGCGGTACAAATGGGAC 688
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 622 CAAGTCCCTTACTCCCAACGCAAGGAGGCTTACCGCGCAATGTTATCTCTTGAAC 680
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||

RESULT 6
B0752115 762 bp mRNA linear EST 18-JUL-2002
LOCUS Colletotrichum trifolii cDNA clone pDSCT10-12, mRNA
DEFINITION sequence.
ACCESSION B0752115
VERSION B0752115.1 GI:21907520
KEYWORDS Colletotrichum trifolii
SOURCE Colletotrichum trifolii
ORGANISM Colletotrichum trifolii
REFERENCE Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
AUTHORS Sordariomycetes incertae sedis; Phyllachorales; Phyllachoraceae;
TITLE mitosporic Phyllachoraceae; Colletotrichum.
JOURNAL 1 (bases 1 to 762)
COMMENT Cheung,F. and Fraser,C.M.
ESTs from mycelia of Colletotrichum trifolii race 1
Unpublished Other_ESTs: EST632677
Contact: Deborah A. Samac
Department of Plant Pathology
University of Minnesota
495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA
Tel: 612 625 1243
Fax: 651 649 5058
Email: debbys@puccini.crl.umn.edu
TIGR sequence name: MSAJ12TV More information is available at:
www.medicado.org
Seq primer: (gtA ATA CgA CtC ACT ATA ggg C).

```

```

FEATURES
source
Location/Qualifiers
1..762
/organism="Colletotrichum trifolii"
/mol_type="mRNA"
/strain="race 1"
/db_xref="taxon:5466"
/clone="pDSCT10-12"
/tissue_type="mycelia"
/dev_stage="young, actively growing mycelia (3 days after
inoculation) grown in liquid culture (cutin minimal medium
containing 2%glucose)."
/lab_host="DH5alpha"
/clone_lib="DSCT"
/notes="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:
EcoRI; Isolate: 2sp2; cDNA was prepared from polyA+
enriched RNA The cDNA was ligated into Lambda gtl1 from
Stratagene and packaged using Gigapack packaging extracts.
An aliquot of the amplified library was used to transduce
E. coli Y1090 and phage DNA was purified from a liquid
lysate. The cDNA inserts were gel purified after EcoRI
digestion and ligated into pBluescript SK+. Aliquots of
the ligation were used to transform E. coli DH5alpha which
were plated onto medium with X-gal for selection of
recombinants."
BASE COUNT 163 a 267 c 221 g 111 t
ORIGIN
Query Match 11.5%; Score 256.4; DB 13; Length 762;
Best Local Similarity 66.4%; Pred. No. 6.8e-64;
Matches 368; Conservative 0; Mismatches 186; Indels 0; Gaps 0;
QY 79 CTTAGCAAGTGCAGGTGGCGCACCAAAACAGGATTTGGTGGCCCAACATGCTCAAC 138
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 114 CTGATGAACACGGCGCGCGCATCAAGAACCGGACTGGTGGCGGATGGCTGAAG 173
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 139 CTCGGCATCTAGCCCAACATTCATCGCTATCGGACCCCAACACCGGATTTTGTACTAT 198
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 174 CTCACATCTCTCGCCAGCAGCACAGCCGCTCACCAACCCGCTGGCGGAGACTTTGACTAC 233
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 199 GCGGAAGAGTAAAGAGCTAGATCTGGCAGCGGTTAAAAAGGACCTGGCAGCGCTAATG 258
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 234 GTCGCGCTTCAAGTCGCTCGACTACGAGGCGCTCAAGAGAGCACTCAGGCGCTGATG 293
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 259 ACAGATTCACAGGACTGGTGGCCAGCAGATTTAGCGTCATTTATGGCCCCCTTTCTTTATACGC 318
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 294 ACCGACTCCAGGACTGGTGGCTTGGCCACTTTGGCCACTACGCGGCTGTGTTTCATCCGC 353
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 319 ATGGCGTGGCAGCGCGCGCACTACCGTATCGGTATGGCGCGTGGTGGCGTGGCTCC 378
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 354 ATGGCCTGGCAGCGCGCGCGCACTACCGAGTTACAGCGAGCGAGGTGGTGGAGAG 413
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 379 GGCTACAGCGCTTTCGCGCTCTCAATAGCTGGCCAGACAATGCCAATCTGGATAAAGCA 438
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 414 GGCCACCAAGCTTTCGACCGCTCAACAGCTGGCCGAGCAATCTACGCTCGACAAGGCC 473
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 439 CGTTCGCTCTTTGGCCCATCAACAAATAATACGTCGAAATAATCTCTGGGGCGGATCTA 498
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 474 CGTCGCTGCTGTGGCCCATCAAGCAAAAGTAGTGGCAACAAAGATCTCTGTGGCGGACCTC 533
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 499 ATGATACTCACAGGAACCTAGCTCTGGAAACTATGGCTTTTAAACTTTTGTGTTTGA 558
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 534 ATGATCTGGCGGCAACGTTGGCCCTCGAGTCCATGGGTTTCCAGACGGCGGCTTCTCC 593
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 559 GGTGGCAGCAGCATGTATGGAGCGCTGAAGAAGATGTATCTACCTGGGAGCAGCAAAACGAA 618
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 594 GGAGGCGCTGGCAGACCTGGGAAGCCGACGAGTCCGCTCTACTTGGGCGCGCAGACACC 653
   || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 619 TGGCTGGGAGACAA 632
   || ||||| || |||||
Db 654 TGGCTGGGACAA 667
   || ||||| || |||||

```

RESULT 7

LOCUS	BQ751801	707 bp	linear	EST 18-JUL-2002
DEFINITION	EST632364 D5CT Colletotrichum trifolii cDNA clone pDSCT7-68, mRNA sequence.			
ACCESSION	BQ751801			
VERSION	BQ751801.1	GI:21907206		
KEYWORDS	EST.			
SOURCE	Colletotrichum trifolii			
ORGANISM	Colletotrichum trifolii			
REFERENCE	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariomycetes incertae sedis; Phyllachorales; Phyllachoraceae; mitosporic Phyllachoraceae; Colletotrichum.			
AUTHORS	1 (bases 1 to 707)			
TITLE	Samac,D.A., Dickman,M., Town,C.D., Van Aken,S., Utterback,T., Cheung,F. and Fraser,C.M.			
JOURNAL	ESTs from mycelia of Colletotrichum trifolii race 1			
COMMENT	Unpublished			
	Contact: Deborah A. Samac			
	Department of Plant Pathology			
	University of Minnesota			
	495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA			
	Tel: 612 625 1243			
	Fax: 651 649 5058			
	Email: debbys@puccini.crl.umn.edu			
	TIGR sequence name: M5AG688v More information is available at:			
	www.medicago.org			
	Seq primer: (gta ata cga ctc act ata ggg c).			

FEATURES

```

location/Qualifiers
1. 707
/organism="Colletotrichum trifolii"
/mol_type="mRNA"
/strain="race 1"
/db_xref="taxon:5466"
/clone="pDSCT7-68"
/tissue_type="mycelia"
/dev_stage="young, actively growing mycelia (3 days after
inoculation) grown in liquid culture (cutin minimal medium
containing 2%glucose)."
/lab_host="DH5alpha"
/clone_lib="DSCT"
/notes="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:
EcoRI; isolate: 25p2 ; cDNA was prepared from polyA+
enriched RNA The cDNA was ligated into Lambda gt10 from
Stratagene and packaged using Gigapack packaging extracts.
An aliquot of the amplified library was used to transduce.
E. coli Y1090 and phage DNA was purified from a liquid
lysate. The cDNA inserts were gel purified after EcoRI
digestion and ligated into pBluescript SK+. Aliquots of
the ligation were used to transform E. coli DH5alpha which
were plated onto medium with X-gal for selection of
recombinants."
146 a 247 c 195 g 119 t

```

BASE COUNT	ORIGIN
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

Query Match	11.4%	Score	254.8	DB 13	Length	707
Best Local Similarity	66.2%	Pred. No.	1.9e-63			
Matches	367	Conservative	0	Mismatches	187	Indels
						Gaps
						0
QY	79	CTTAAGCAAAGTGCAGGTGGCGGCACCAAAAACAGGATTGTCGCCCAACATGCTCAAC	138			
Db	140	CTGATGACACCGGCGGCGGCATCAAGAACCGGACTGTGGCCGGATCGCGTGAAG	199			
QY	139	CTTCGGCATCTTACGCCAACATTTATCGTATCGGACCCCAACAGCCCGGATTTTGACTAT	198			
Db	200	CTCAACATCTCTCGCGACACACGCCCGCTACCAACCCGCTGGCGCAGGACTTTCCTAC	259			
QY	199	GCCGAAGAGTTTAAAGACTAGACTGGCAGCGGTTAAAGAGACTGSCACGCGTAAAG	258			
Db	260	GTCGCCGCTTCAAGTCGCTGCATACGAGGCGCTCAAGAGGACTCACGGCCCTGATG	319			
QY	259	ACAGATTACAGGACTGTTGGCCACGAGATTACGGTCAATTATGGCCCCCTCTTTTATACGC	318			
Db	320	ACCGACTCCAGAGCTGTGGCTCCGCACTTTGGCCACTACGGGGGTCTGTTTCATCCGC	379			

319	ATGGCGTGGCAGCGCCGGACCTACCGTATCGGTGATGGCGTGGTGGCGTGGCTCC	378
380	ATGGCGTGGCAGCGCCGGACCTACCGAGTTACGACGCGCGAGGTGGTGAGAG	439
379	GGCTCACAGCGGTTGCGCCCTCTCANTAGCTGGCCACAGAAATGCCAAATCTGGATAAAGCA	438
440	GGCCAGCAGCGTTGCGCACCGCTCACAGAGCTGGCGGACAAATGTACGCTTCGACAAGGCC	499
439	CGCTTGCCTTTTGGCCCATCAACAACAAATACGGTCGAAAAATCTCTGGGGCGGATCTA	498
500	CGTCGGCTGCTGTGGCCCATCAGCAAAAGTACGGCAACAAGATCTCGTGGGCGCGACTTC	559
499	ATGATACTCAGAGAAAGTAGCTCTGGAAACTATGGCTTTTAAAACTTTTGGTTTGGCA	558
560	ATGATCTTGGCGGCAAGCTGGCCCTCGAGTCCATGGGTTTCAGACAGGCCCGGCTTCTCC	619
559	GGTGGCAGACAGATGTATGGGAGCCCTGAAGAAGATGTATCTGGGGAGCAGCAAAACCGAA	618
620	GGAAAGCCGTCCGCACACCTGGGAAGCCGACAGAGTCCGCTCTACTGGGGCGCGAAGAACAC	679
619	TGGCTGGGAGACAA	632
680	TGGCTCGGCAACAA	693

[illegible]

```

FEATURES
source
Plate: / row: G Column: 22.
Location/Qualifiers
1. 710
/organism="Gibberella zeae"
/mol_type="mRNA"
/strain="NRRL 31084"
/db_xref="taxon:5518"
/clone_lib="Nitrogen-starved mycelia"
/note="Vector: Uni-ZapII; Site_1: EcoRI; Site_2: XhoI"
139 a 238 c 192 g 141 t
BASE COUNT

```

Query Match	11.18;	Score 249.4;	DB 13;	Length 710;
Best local Similarity	60.8%;	Pred. No. 7.6e-62;		
Matches 425;	Conservative 0;	Mismatches 271;	Indels 3;	Gaps 17;
1480	CGCGCGGTGCCACGGTGCACGTATACGACTGGCCGCCCAAAAAAGACTGGGAAGTAAC	1539		
12	CGTGGTGGTCCACGGTGGCCGTCTCCGCTCTCGAGCCCCAGAAAGAACTGGGAGGTCAAC	71		
1540	AACCGCTCAGCAACTTGCAGGGGTACTCAAAACACTAGGAAGGTATCCAGGAGGACTTTTAAAC	1599		

```

Db      72  AACCCTCTCAGCTGGGCGAGGTTCTCAACGGTCTTGAGAAGATCCAGAAACATTCAC 131
QY      1600 GAGGCGCAATCAGATAACAAGACAGTATGTTGGCGCACTGATGTGTGTCGGCGGTGT 1659
Db      132  GACTCTCAGTCCACGGCAAGGCCATCTCTCTTGCTGACGTGATGCTGCTGCTGTTCT 191
QY      1660 GCGGTGTAGAAAAAGCTGCANAAAGATGCTGCGCATGAGGTGCGAGTGTTCACCGG 1719
Db      192  GTCGCGGTGCGAAGAGCGCGCAAGATGCGGTGTTAAGCTCAGCTGCGCTTCACTCTCT 251
QY      1720 GGACGAGCGGATGCGCACCGCTGAGCAAAACCGATGGAAGCTTTCGAAGCACTAGAGCA 1779
Db      252  GGCCTGACGATGCCACCGACGAGCAGCAGCGAGTCCAGTCTGTGCTCACCTCGAGGCC 311
QY      1780 GCGGTGACGGCTTTAGAAACTACATTAACCGGAGCATAAAGTATCCGCTCAGGAATG 1839
Db      312  GCGCGGACGGTTTCGAAATTAACGGCGCTTCCACAGACCGGTGTCGAAGTCCAGCACATG 371
QY      1840 CTCGTAGACCGGCGCAGCTTCTGCTGCTTTCGGCACCAGAAATGACTGCTTGGTAGGC 1899
Db      372  CTCATCGACCGCGCCAGCTTCTCACTCTCAGCTGCCCGAGATCACCGCTGCTGCTGCTG 431
QY      1900 GGTATGCGGTACTGGGCACCAACTAGCAGCGTTCGCAGCATGGAGTGTTCACAAATAG 1959
Db      432  GGTCTGTGCTCTCAACCGCAACTGGGACGGATCATCCACGGTATCTCCACAGGCGC 491
QY      1960 CCGGTGACGTATCCAACTGATCTTGTGTAACCTGCTAGACCTCAACACTAAATGGCGA 2019
Db      492  CTGTGTGTTCTGTCACACGACTTCTGTCACCTCTCGACATCAGCACCGAGTGGAG 551
QY      2020 GCCAGCGATGATCAGACAAAGTTTGTGAGCGACAGACTTCAAACTGGCGAAGTAAAG 2079
Db      552  CCGCTGCGCAA---CGCGCACTGTCGAGGGTATTGACCGCAAGACGCGCAACAAGAG 608
QY      2080 TGGAGTGGCACCGCGGTAGACCTGATCTTCGGATCCAAATTCGGAGCTAAGACCTCGCA 2139
Db      609  TGGACTGTACCGGTGTCGATCTGCTTGTGATCAGCTGAGCTCGCGCGACAGCT 668
QY      2140 GAAGTGTAGGCTGTGCAGATCTTGAAGAAAAAGTTGTT 2178
Db      669  GAGACATATGCTGAAGCTGGCGGTAGCGGAGAGCTGGTT.707

RESULT 9
AF075904
LOCUS   AF075904 Salmonella typhimurium LT2, Lambda DASH II Salmonella
DEFINITION
trophimurium genomic clone B173-T3, genomic survey sequence.
ACCESSION
AF075904
VERSION
AF075904.1 GI:3320774
KEYWORDS
GSS.
SOURCE
Salmonella typhimurium
ORGANISM
Salmonella typhimurium
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
REFERENCE
1 (bases 1 to 918)
Wong,R.M.-Y., Wong,K.K., Benson,N.R. and McClelland,M.
Sample sequencing of a Salmonella typhimurium LT2 lambda library:
comparison to the Escherichia coli K12 genome
FEMS Microbiol. Lett. 173 (2), 411-423 (1999)
JOURNAL
9243757
MEDLINE
10227170
PUBMED
10227170
COMMENT
Contact: McClelland M
Molecular Biology
Sidney Kimmel Cancer Center
3099 Science Park Road, San Diego, CA 92121, USA
Email: mclelland@lifsci.sdsu.edu
Class: shotgun.
Location/Qualifiers
1..918
/organism="Salmonella typhimurium"
/mol_type="genomic DNA"

FEATURES
source

```

```

/strain="LT2"
/db_xref="taxon:602"
/clone="B173-T3"
/clone_lib="Salmonella typhimurium LT2, Lambda DASH II"
/note="Vector: Lambda DASH II; sequenced using Li-Cor
sequencer."
BASE COUNT      188 a      248 c      283 g      199 t
ORIGIN
Query Match      10.9%; Score 244.6; DB 28; Length 918;
Best Local Similarity 58.9%; Pred. No. 2.3e-60;
Matches 532; Conservative 0; Mismatches 334; Indels 37; Gaps 5;
QY      900  GAGCCTGGGTGGAAAAACACCTACGGCACCGACACAGTGGGGATACCATCACCAGTGG 959
Db      24   GGGCTTAGTGTGGCGCAGCAGCTATGTTAGTGGCGTGGCGGGATGCTATCACCTCCGG 83
QY      960  ACTAGAAGCGCTGTGACCAAGACCCCTACTCAATGGAGCAATACTTTTGGAAAACT 1019
Db      84   GCTGGAATGGTCTGGACGACAGCGCCAGTGGGCAACTATTTCTTCGAGAACCT 143
QY      1020 CTTTGGTTACGAGTGGGAGCTTACCAAAAGTCCAGCTGGAGCTTATCATGTGAAACCAA 1079
Db      144  GTTCAATATGATGGGTACAAACCGTAGTCCGGCTCGGCTATCCAGTTTGAACGGT 203
QY      1080 AGACGGTGGCGGGGTGGCACCATACCGGATGACATGCCAGCAATGCCAGCAAGTCGACGCTCC 1139
Db      204  AGA-----CGCGCGGACATCATCCGACCGCTTCGATCCGTCGAAAAAACGTAAGCC 257
QY      1140 ATTTATGCTCACTACGGACCTGGCGCTGGCATGGACCTGATAGCAAAAATTTCTCG 1199
Db      258  AACCATGTGCTACCGGCTGACGCTGCTTTGATCCGGAGTTCGAGAAGATTTCCCG 317
QY      1200 ACGGTACTATGAAACCCCTGATGAGTTTGCAGATGTTTCGGAAGACATGTTACAACT 1259
Db      318  TCGTTTCTTAACGATCC--GCAGCCTTTAATGAAGCTTTGCTGCTGCTGTTCAACT 376
QY      1260 GACACACAGATATGGGACCAAGGTGCGCTACCTGGGACCAAGAGTGCCTCAGGAAGA 1319
Db      377  GACGCACAGATATGGGACCAAAAGCGGTTACATCGGACCGGAGTCCGGAAGAAGA 436
QY      1320 CCTCATCTGCGAAGACCTTATACAGATGTAAGCCATCTCTGTAGACGAAACGATAT 1379
Db      437  TGTGATCTGGCAGGACCCGTTGCCGCAACCGCTCTATCAGCCACCCAGGAGACATAT 496
QY      1380 TGAAGCCTTAAAGCCAAAATCCTGGAATCGGAGCTGACGGTAAAGCGAGCTGTAAGCAC 1439
Db      497  CAA---CCTGAAAGCGCGATCGTGCATCCGGGCTTCTATTAGCGAGATGTTTCGGT 553
QY      1440 GGCATGGGCTTCTGCATCTACTTTAGAAACTCTGACAAGCGCGGGGTGCCAACGGTGC 1499
Db      554  TGCCTGGGATCCGCTCTACTTTCGCGCGCGCGGATAAGCGTGGCGGCTAAGCGCGC 613
QY      1500 AGGTATACACTGGCCCCACAAAAGACTGGGAAGTAAACACCCCTCAGCAACTTCCAG 1559
Db      614  GGTCTGGCATATTAGCCCTCAGCGCATGGGATGTCAACGCGGCAATCAGATAACAA 1619
QY      1560 GGTACTCAAAACACTAGAAGGTATCAGGAGGACTTTAAACGAGCGCAATCAGATAACAA 1619
Db      668  CGTCTCGCGGTATTAGAAGAGATCCAGAA-----AACGAGGAA 706
QY      1620 AGCAGTATCGTTGGCGGACCTGATTTGCTGCGCGGCTGTGCGGGTGTAGAAAAAGTGC 1679
Db      707  TAAAGCTCGCTGGCGGATATTATTGCTGCGCGGCGGTGGTGGTATCGACGAGCGGC 766
QY      1680 AAAAGATGCTGGCCATGAGTGCAGTGCCTTTCACCCGGGACGAGCGGATCCACCGC 1739
Db      767  CGCTGCTGGGGTGTGAGCATCAGCTACCTTTTTCGCCGCGCGGTGAGCGGTCA 836
QY      1740 TGAGCAAAACCGATGTGGAAGCTTTTCGAAGCACTAGAGCAGCGGCTGACGGCTTTAGAA 1799
Db      827  GGATCAGACCGACATCGAGATGTTCTCGCTGCTTGAACCGATGCTGATGGATTCGGTAA 886

```

```

QY 1800 CTA 1802
Db 897 CTA 889

RESULT 10
CNS01HHH 737 bp DNA linear GSS 01-JUN-2001
LOCUS Anopheles gambiae GSS sp6 end of clone 08M05 of NotreDamel library
DEFINITION from strain PEST of Anopheles gambiae (African malaria mosquito),
genomic survey sequence.
AL144438
AL144438.1 GI:7002600
GSS.
Anopheles gambiae (African malaria mosquito)
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae,
Anopheles.
1 (bases 1 to 737)
REFERENCE Genoscope.
AUTHORS Direct Submission
TITLE Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
2 (bases 1 to 737)
REFERENCE Roth,C.W., Brey,P.T., Ke, Z., Collins,F.H. and Weissenbach,J.
AUTHORS Direct Submission
TITLE Submitted (16-FEB-2000) BMMI, Institut Pasteur, 25, rue du Dr.
JOURNAL Roux, Paris 75015, France
COMMENT This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.
FEATURES
source Location/Qualifiers
1..737
/organism="Anopheles gambiae"
/mol_type="genomic DNA"
/strain="PEST"
/db_xref="taxon:7165"
/clone="08M05"
/clone_lib="NotreDamel"
/note="end : SP6"
BASE COUNT 161 a 217 c 224 g 127 t 8 others
ORIGIN
Query Match 10.9%; Score 243; DB 29; Length 737;
Best Local Similarity 62.2%; Pred. No. 6e-60;
Matches 457; Conservative 7; Mismatches 257; Indels 14; Gaps 5;

QY 341 CCTACCTATCGGTGATGGCGGTGGTGGCGGTCCGGCTCACAGCGCTTCGCGCTC 400
Db 16 CATATCGTCCGTGGATGGCGGTGGCGGAGCAGCGGTCAACAGCGCTT--TGCCAC 73

QY 401 TCATAGCTGCCAGACAAATGCCAATCTGTGAATGACGACGCTGCTTCTTTGGCCATCA 460
Db 74 TGAATCTCTGGCGGATGAACGTAAAGCTGTGATAAACGCGCGCTCTCTGTGGCCTATCA 133

QY 461 AACAAAAATACGGTCGAAAAATCTCCCTGGCGGATCTAATGATACTCACAGGAACCTAG 520
Db 134 AGAAAAATACGACCAAAAAATCTCTGGCGGACCTGTTATCTCTCGCGGTAAACGTGG 193

QY 521 CTCTGGAAATATGGCGCTTTAAACTTTTGGTTTGGAGTGGCGAGCAGATGATATGG 580
Db 194 CGTGGAGAACTCCGGCTTCCGTACCTTTTGGTTTGGTGGCGGCGTGAAGACGCTCTGG 253

QY 581 AGCTGAAGAAGATGATCTACTGGGAGCAGAAACCGAATGCTGGGGAGACAGCGCTATG 640
Db 254 AACCGGATCTTGAGTGAATGGGGTATGAAAAAACCTGGCTG-----AMCACCAGCG 307

QY 641 AAGGTGACGAGCGCTGCGAAATCCCTGGGAGCGCTTACAAATGGGACTCATCTATGTA 700
Db 308 ACCCGAAGCGCTGGCTAARCGTCCCTCTGGCAGCCACCGAATGGGCTGATCTACGTGA 367

```

```

QY 701 ACCCGAAGGACCCAAACGCAAGCCAGACACCTATCGCTGCTGCGGTGATATTCGTGAGA 760
Db 368 ACCCGAAGG--CCGAACGCCAGCGGTGAACCGCTGTGAGCGCGGCACCGATTCTGTGCGA 425

QY 761 CTTTGGCCGAATGCAATGACGAAGAACCGTGGCTCTCATAGCGGTGGACACA 820
Db 426 CTTTGGCAACATGGCATGAACGATGAAGAGACCGTGGCTGATTCAGCGCGGCCACA 485

QY 821 CTTTGGAAAAACCATGGTCTGCCGATGCGGAGAAATATGTGGGCGGAGAGCGCTGCCG 880
Db 486 CCCTGGCAAGACTCACGGCGCGGTGAAGCC--ACCCACGTGGGTACCGACCCCGAAG 542

QY 881 CCGCAGGTATTGAAGAAATGAGCTGGGTGGAAAAACACACCTACGGCACCGGACACGGTG 940
Db 543 CCTCACCGATTAAAMRCACAGCGCTGGCTGGCCAGCAGCCAGCGGCGGTGGCG 602

QY 941 CGGATACCATCACCAGTGGACTAGA-AGGCGCGCTGGAGCAAGACCCCTACTCAATGGAGC 999
Db 603 CTGATGCCATTACCTCCGGTCTGGACAGTCTCTGATCAGTACCAAAACCCCGACCGTGGAGC 662

QY 1000 AATAACTTTTGTAAACCTCTTGGTTACGAGTGGGAGCTTACCAAAAGTCCAGCTGGA 1059
Db 663 AACTACTTCTTCGAGAACCTGTTCAAATACGAATGGGTACAGACCCGACCCGCGCAGGS 722

QY 1060 GCTTATCAGTGGAAA 1074
Db 723 GCGATTCAAGTTTGA 737

```

RESULT 11

BQ751881

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1..742

/organism="Colletotrichum trifolii"

/mol_type="mrna"

/strain="race 1"

/db_xref="taxon:5466"

/clone="pDSC19-27"

/tissue_type="mycelia"

/dev_stage="Young, actively growing mycelia (3 days after inoculation) grown in liquid culture (cutin minimal medium containing 2%glucose)."

/lab_host="DH5alpha"

/clone_lib="DSC1"

/note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:

742 bp mRNA linear EST 18-JUL-2002

ES7632444 DSCT Colletotrichum trifolii cDNA clone pDSC78-27, mRNA

sequence.

BQ751881

EST

Colletotrichum trifolii

Colletotrichum trifolii

Eukaryota; Fungi; Ascomycota; Perizomycotina; Sordariomycetes;

Sordariomycetes incertae sedis; Phyllachorales; Phyllachoraceae;

Mitosporic Phyllachoraceae; Colletotrichum.

1 (bases 1 to 742)

Samad,D.A., Dickman,M., Town,C.D., Van Aken,S., Utterback,T.,

Cheung,F. and Fraser,C.M.

ESTs from mycelia of Colletotrichum trifolii race 1

Unpublished

Other_ESTs: EST632443

Contact: Deborah A. Samac

Department of Plant Pathology

University of Minnesota

495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA

Tel: 612 625 1243

Fax: 651 649 5058

Email: debby@puccini.crl.umn.edu

TIGR sequence name: M7SAH27TV More information is available at:

www.medicago.org

Seq primer: (gtA AtA CgA Ctc Act AtA ggg C).

Location/Qualifiers

1..742

/organism="Colletotrichum trifolii"

/mol_type="mrna"

/strain="race 1"

/db_xref="taxon:5466"

/clone="pDSC19-27"

/tissue_type="mycelia"

/dev_stage="Young, actively growing mycelia (3 days after inoculation) grown in liquid culture (cutin minimal medium containing 2%glucose)."

/lab_host="DH5alpha"

/clone_lib="DSC1"

/note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:

[illegible]

[illegible]

Search completed: October 8, 2003, 15:29:20
Job time : 4640.63 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 7, 2003, 19:31:22 ; Search time 60.1769 Seconds
(without alignments)
1965.061 Million cell updates/sec

Title: US-09-884-889-8
Perfect score: 4002
Sequence: 1 MENKHGSSSYNTNGKCK.....VKDFVKAQKVMDLDRFDLK 745

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4002	100.0	745	19 AAW33810	Microscilla furves
2	2544	63.6	740	18 AAW26596	Mycobacterium tube
3	2544	63.6	740	20 AAW31343	Mycobacterium tube
4	2544	63.6	740	20 AAW95398	M. tuberculosis ca
5	2541	63.5	740	20 AAW95399	M. tuberculosis ca
6	2290.5	57.2	753	19 AAW33809	Alcaligenes (Deley
7	2290.5	57.2	753	23 AAW33809	Alcaligenes (Deley
8	2278.5	56.9	735	14 AAW33670	M.tuberculosis H37
9	2278.5	56.9	735	20 AAW78363	M.tuberculosis kat

10	2201.5	55.0	723	23	ABB78212	Amino acid sequenc
11	1699.5	42.5	731	11	AA080205	Heat resistant per
12	1366	31.6	533	22	ABG25057	Novel human diago
13	883.5	22.1	273	22	ABG08823	Novel human diago
14	675	16.9	1550	22	ABG09297	Novel human diago
15	455	11.4	201	22	ABG08828	Novel human diago
16	278.5	7.0	279	23	ABB93261	Herbicidally activ
17	273.5	6.8	291	23	ABB05661	Herbicidally activ
18	266.5	6.7	246	23	ABB92257	Herbicidally activ
19	246	6.1	297	21	AAG26452	Arabidopsis thalia
20	246	6.1	297	21	AAG45588	Arabidopsis thalia
21	244	6.1	250	21	AAG03702	Arabidopsis thalia
22	244	6.1	250	21	AAG16805	Arabidopsis thalia
23	244	6.1	250	21	AAG26453	Arabidopsis thalia
24	244	6.1	250	21	AAG43523	Arabidopsis thalia
25	244	6.1	250	21	AAG45582	Arabidopsis thalia
26	244	6.1	250	21	AAG45585	Arabidopsis thalia
27	244	6.1	250	21	AAG45589	Arabidopsis thalia
28	244	6.1	250	23	ABH90903	Herbicidally activ
29	244	6.1	254	21	AAG43522	Arabidopsis thalia
30	240.5	6.0	287	21	AAG09460	Arabidopsis thalia
31	240.5	6.0	287	21	AAG43316	Arabidopsis thalia
32	240.5	6.0	287	23	ABH93244	Herbicidally activ
33	232	5.8	215	21	AAG09703	Arabidopsis thalia
34	232	5.8	215	21	AAG16806	Arabidopsis thalia
35	232	5.8	215	21	AAG26454	Arabidopsis thalia
36	232	5.8	215	21	AAG43524	Arabidopsis thalia
37	232	5.8	215	21	AAG45583	Arabidopsis thalia
38	232	5.8	215	21	AAG45586	Arabidopsis thalia
39	232	5.8	215	21	AAG45590	Arabidopsis thalia
40	232	5.8	354	21	AAG44544	Arabidopsis thalia
41	232	5.8	426	21	AAG44543	Arabidopsis thalia
42	232	5.8	426	23	ABH91634	Herbicidally activ
43	232	5.8	452	21	AAG44542	Arabidopsis thalia
44	229.5	5.7	287	21	AAW77933	A. thaliana enviro
45	223.5	5.6	254	21	AAG09461	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAW33810
ID AAW33810 standard; Protein; 745 AA.
XX
AC AAW33810;
XX
DT 18-NOV-1998 (first entry)
XX
DE Microscilla furvescens catalase-53A1.
XX
KW Catalase; epoxidation; hydroxylation; biosensor; paper bleaching;
KW pasteurisation.
XX
OS Microscilla furvescens.
XX
PN WO9800526-A1.
XX
PD 08-JAN-1998.
XX
PF 03-JUL-1997; 97WO-US16513.
XX
PR 03-JUL-1996; 96US-0674887.
XX
PA (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX
PI Adhikary RS, Robertson DE, Sanyal I;
XX
DR WPI: 1998-086953/08.
XX
DR N-PSDB; AAV06555.
XX
PT New bacterial catalases, related nucleic acid vectors and
PT transformed cells - used as oxidising agents and for detecting or

PT destroying hydrogen peroxide, e.g. in biosensors
 PS Claim 1; Fig 2; 35pp; English.
 XX

CC The present sequence is of the *Microscilla furvescens* catalase-53CA1.
 CC Catalase-53CA1 may be used to catalyze oxidation reactions such as
 CC epoxidation or hydroxylation. The enzyme can also be used to detect or
 CC destroy hydrogen peroxide, e.g. in connection with glyoxylic acid
 CC production, biosensors, contact lens cleaning, pulp/paper bleaching and
 CC pasteurisation of dairy products. Antibodies raised against
 CC catalase-53CA1 can be used to screen libraries for detection and
 CC purification of cells containing the enzyme. Fragments of the gene
 CC encoding Catalase-53CA1 can be used to identify related sequences.
 XX
 SQ Sequence 745 AA;

Query Match 100.0%; Score 4002; DB 19; Length 745;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 745; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MENKHSGSSTYNTNTGKCPFTGGSLKQSGAGGTKNRDWWPNMLNLGILRQHSLSLSDPN 60
 Db 1 MENKHSGSSTYNTNTGKCPFTGGSLKQSGAGGTKNRDWWPNMLNLGILRQHSLSLSDPN 60

Qy 61 DPDPDYAEERKKDLAAVKKDLAALMTDSQDWPADYGHYGFPIRMAHWSAGTYRIGDG 120
 Db 61 DPDPDYAEERKKDLAAVKKDLAALMTDSQDWPADYGHYGFPIRMAHWSAGTYRIGDG 120

Qy 121 RGGGSGSRFAPLNSPNDANLDKARLLLPWKQYGRKISWADLMILGNVALETWGF 180
 Db 121 RGGGSGSRFAPLNSPNDANLDKARLLLPWKQYGRKISWADLMILGNVALETWGF 180

Qy 181 KTEFGAGRADWPEEDVYWGAEPLGDKRYEGDRELENPLGAVQMGILYVNPPEGNG 240
 Db 181 KTEFGAGRADWPEEDVYWGAEPLGDKRYEGDRELENPLGAVQMGILYVNPPEGNG 240

Qy 241 KPOPIAARDIRETFGRMANDEETVALIAGGHTFGTHGAADAERKYGREPAAGIEEM 300
 Db 241 KPOPIAARDIRETFGRMANDEETVALIAGGHTFGTHGAADAERKYGREPAAGIEEM 300

Qy 301 SLGKNTYGTGCHGADTTISLEGAWTKTPTQWSNFFENLFGYEWELTKSPAGAYQWKPK 360
 Db 301 SLGKNTYGTGCHGADTTISLEGAWTKTPTQWSNFFENLFGYEWELTKSPAGAYQWKPK 360

Qy 361 DGAGAGTIPDAHDPKSHAPFMTLTDIALRMDPDYKISRYYENPDEFADAFKAWYKL 420
 Db 361 DGAGAGTIPDAHDPKSHAPFMTLTDIALRMDPDYKISRYYENPDEFADAFKAWYKL 420

Qy 421 THRDMPKRYLGPVPEQDLIWODPIPDVSHPLVDENDIEGLKAKLTESGLTVSELVST 480
 Db 421 THRDMPKRYLGPVPEQDLIWODPIPDVSHPLVDENDIEGLKAKLTESGLTVSELVST 480

Qy 481 AWASASTFRNSDKRGANGARIRLAPQKDEWVNNPQOLARVLKTLGIGEDFNQAOQSDNK 540
 Db 481 AWASASTFRNSDKRGANGARIRLAPQKDEWVNNPQOLARVLKTLGIGEDFNQAOQSDNK 540

Qy 541 AVSLADLIVLAGCAGVEKAAKAGHEVQVPFNPGRADATAEQTDVEAFEALPAADGFRN 600
 Db 541 AVSLADLIVLAGCAGVEKAAKAGHEVQVPFNPGRADATAEQTDVEAFEALPAADGFRN 600

Qy 601 YIRPEHKVSAEMLVDRQAQLLSAPEMTALVGMRLVLTNYDGSQHGVTNRPQGLSND 660
 Db 601 YIRPEHKVSAEMLVDRQAQLLSAPEMTALVGMRLVLTNYDGSQHGVTNRPQGLSND 660

Qy 661 FFVNLDLNTKWRASDESDKVFEGRDPKTGEVWKSGRVPLIFGNSSELRALAEVYGCAD 720
 Db 661 FFVNLDLNTKWRASDESDKVFEGRDPKTGEVWKSGRVPLIFGNSSELRALAEVYGCAD 720

Qy 721 SEEKFKVDFKAWKVMDLDRFDLK 745
 Db 721 SEEKFKVDFKAWKVMDLDRFDLK 745

RESULT 2
 AAW26596
 ID AAW26596 standard; Protein; 740 AA.
 XX
 AC AAW26596;
 XX
 DT 25-MAR-2003 (updated)
 DT 21-JAN-1998 (first entry)
 XX
 DE Mycobacterium tuberculosis wild-type catalase-peroxidase.
 XX
 KW Isoniazid resistance; restriction fragment length polymorphism;
 KW RFP; katG gene; catalase; peroxidase; multiple drug resistance;
 KW isonicotinic acid hydrazide; INH; tuberculosis; antibiotic;
 KW tuberculostatic.
 XX
 OS Mycobacterium tuberculosis.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 264 /note= "mutation to Thr confers INH resistance"
 FT Misc-difference 315 /note= "mutation to Thr confers INH resistance"
 FT Misc-difference 337 /note= "mutation to Cys confers INH resistance"
 FT Misc-difference 463 /note= "mutation to Leu confers INH resistance"
 XX
 US5658733-A.
 XX
 PD 19-AUG-1997.
 XX
 XX 07-APR-1995; 95US-0418782.
 XX
 XX 07-APR-1995; 95US-0418782.
 PR 18-APR-1994; 94US-0228662.
 XX
 XX (MAYO-) MAYO FOUNDATION.
 PA
 XX
 PI Cockerill FR, Kline BC, Uhl JR;
 XX
 DR WPI; 1997-424226/39.
 DR N-PSDB; AAT90400.
 XX
 PT Determining susceptibility of Mycobacterium tuberculosis strains to
 PT isoniazid - by detecting mutation(s) in the catalase-peroxidase gene,
 PT katG
 XX
 PS Example 2; Column 31-36; 38pp; English.
 XX
 CC This polypeptides comprises a consensus of the Mycobacterium
 CC tuberculosis wild-type catalase-peroxidase enzyme. Claimed
 CC methods of rapidly identifying strains of M. tuberculosis which are
 CC resistant to the tuberculostatic drug isoniazid (INH) are based on
 CC the discovery of 4 mutations in the wild-type katG gene sequence
 CC (see AAT90400) that confer INH resistance and which coincidently
 CC result in the addition or deletion of restriction endonuclease
 CC sites. Primers (see AAT90401-02) are provided for the PCR
 CC amplification of a katG gene from a test M. tuberculosis strain,
 CC and further primers (see AAT90403-06) for use in the restriction
 CC fragment length polymorphism analysis of the amplified gene and
 CC hence determination of the susceptibility to INH of the strain.
 CC (Updated on 25-MAR-2003 to correct PF field.)
 XX
 SQ Sequence 740 AA;

Query Match 63.6%; Score 2544; DB 18; Length 740;
 Best Local Similarity 64.3%; Pred. No. 1.4e-215;
 Matches 480; Conservative 88; Mismatches 168; Indels 10; Gaps 8;

Qy 2 ENKHSGSSTYNTNTGKCPFTGGSLKQSGAGGTKNRDWWPNMLNLGILRQHSLSLSDPN 61
 Db 3 EQHPPTTETTTGAASNG-CPVVGHMKYPVEGGG--NODWPNRLNLKLVHQNPAVADPMG 59

QY 62 PDFDYAEFFKLDLAANKKDLAALMTDSQDWWPADYGHYGPFFIRMAWHSAGTYRIGDGR 121
 DB 60 AAFDYAAEATIDVDALTRDIEEVNTTSQWPPADYGHYGPLFIRMAWHAAGTYRIHDDR 119
 QY 122 GGGGSGRFPAPLNSWPNANLKDARLLWPDKYGRKISWADLWLTGNVALETMGFK 181
 DB 120 GGAGGGMORFAPLNSWPNASLDKARLLWPVKYKGLSWADLVFAGNCALSGMGFK 179
 QY 182 TFGFAGGRADVWEPEEDYWGAEETWLGDKRYEGDRELENPLGAVQMGLIIVNPGNGK 241
 DB 180 TFGFGFRVDQWEPDE-VTWGKEATWLGDERYSGRKRDLENPLAAVQMGLIIVNPGNGN 238
 QY 242 PDPIAAARDIRETFGRMAAMDEETVALLTAGHTFGTKGADADAERYVGRPAAGIEEMS 301
 DB 239 PDPMAAAVDIRETFGRMAAMDEVETAAALIVGGHTFGKTHGAGPAD-LVGEPEEAPLEQMG 297
 QY 302 LGWKNTYGTGHGADITTSGLGAWTKTPTQWNNFFENLFGYEWELTKSPAGAWKPKD 361
 DB 298 LGWKSSYGTGKDAITSGIEVVTNTPTKWDNSFLEILYGYEWELTKSPAGAWYTKAD 357
 QY 362 GAGAGTIPDAH-DPSKSHAPFMTTDLALRMDPDYEKISRYYENPDFAFAKAWYKL 420
 DB 358 GAGAGTIPDPFGGPGRS--PTMLATDLSLRVDPIYERITRWLEHPEELADEFAKAWYKL 415
 QY 421 THRDMPKRVYLGPEVPOEDLIWQDPIPDVSHPLVDENDIEGLKAKILESGLTVSELVST 480
 DB 416 THRDMPGVARYLGPLVPKQTLWQDPVPAVSHDLVGEAEIASLSQIRASGLTVSQLVST 475
 QY 481 AWASASTFRNSDKRGANGARIRLAPQKDWENNPQ-OLARVLTKEGIEQEDFNOAQSDN 539
 DB 476 AWAAASSFRGSDKRGANGRRIRLQPVQWENNDPDGDLRKVIRTLERIEQSFNSAAPGN 535
 QY 540 KAVSLADILVLGACAGVEKAADAGHEVOVPFNPGRADATAPQTDVEAFEALEPAADGFR 599
 DB 536 IKVSFADLVVLGGCAAIEKAAKAAGHNTVPTPGRTDASQEQTDVESFAVLEPKADGFR 595
 QY 600 NYIKPEHKVSAEEMLVDRALQLLSAPENTALVGGMRVLTGNYDGSQHGVTNKGQOLSN 659
 DB 596 NYLKGKGNLPAPYMLDKANLLTSLAPENTVLVGGRLVGLGANYKRLPLGVFTFSESELTN 655
 QY 660 DFFVNLDLNTKWRASDESDEKVFEGRDKTKTEGVKWSGTRVDLIFGNSSELRALAVYGA 719
 DB 656 DFFVNLDMGWTWESPADDTYQKGD-GSGKVKWTGSRVDLVFGNSSELRALVEYVGD 714
 QY 720 DSEKFKVDKVKAAKVMMDLDRFDLK 745
 DB 715 DAQPRFVQDFVAAMDKVMNLDREFVR 740

RESULT 3

AAW31343

ID AAW31343 standard; Protein; 740 AA.

AC AAW31343;

XX AAW31343;

XX 20-MAR-1998 (first entry)

XX XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

(MAYO-) MAYO FOUNDATION.

Cockerill FR, Kline BC, Uhl JR;

WPI; 1998-007975/01.

Determination of isoniazid sensitivity of Mycobacterium tuberculosis strains - by restriction length polymorphism analysis of katG gene

Example 2; Column 15-20; 18pp; English.

This protein sequence represents the consensus amino acid sequence of the katG gene from Mycobacterium tuberculosis strains H37Rv MC, ATCC 25618, ATCC 27294, G6108, AAH35827, L6627-92, AAL68372, AAL11150, CC AAL24204, AAL33308, AAL16980, L1781, TMC 306, AAL10373 and AAL23261. This sequence is which is used in a novel method to rapidly identify strains of M. tuberculosis which are resistant to isoniazid (INH, isonicotinic acid hydrazide). The method involves the use of restriction fragment length polymorphism (RFLP) analysis to determine if a NciI-MspI restriction site is absent in the DNA of the strain at the codon corresponding to codon 463 of this katG gene consensus sequence. The absence of the site indicates an INH-resistant strain.

SQ Sequence 740 AA;

Query Match 63.6%; Score 2544; DB 19; Length 740;

Best Local Similarity 64.3%; Pred. No. 1.4e-215;

Matches 480; Conservative 88; Mismatches 168; Indels 10; Gaps 8;

2 ENHKGSSSTYNTTGGKCPPTGSLKQSGAGGTTNRDWPNNMLLGLRQHSLSLDPND 61

3 EGHPPITETTTGAASNG-CPVVGHMKYPVEGG--NQDWPNNRLNKLVLHQNPVADPMG 59

62 PDFDYAEFFKLDLAANKKDLAALMTDSQDWWPADYGHYGPFFIRMAWHSAGTYRIGDGR 121

60 AAFDYAAEATIDVDALTRDIEEVNTTSQWPPADYGHYGPLFIRMAWHAAGTYRIHDDR 119

122 GGGGSGRFPAPLNSWPNANLKDARLLWPDKYGRKISWADLWLTGNVALETMGFK 181

120 GGAGGGMORFAPLNSWPNASLDKARLLWPVKYKGLSWADLVFAGNCALSGMGFK 179

182 TFGFAGGRADVWEPEEDYWGAEETWLGDKRYEGDRELENPLGAVQMGLIIVNPGNGK 241

180 TFGFGFRVDQWEPDE-VTWGKEATWLGDERYSGRKRDLENPLAAVQMGLIIVNPGNGN 238

242 PDPIAAARDIRETFGRMAAMDEETVALLTAGHTFGTKGADADAERYVGRPAAGIEEMS 301

239 PDPMAAAVDIRETFGRMAAMDEVETAAALIVGGHTFGKTHGAGPAD-LVGEPEEAPLEQMG 297

302 LGWKNTYGTGHGADITTSGLGAWTKTPTQWNNFFENLFGYEWELTKSPAGAWKPKD 361

298 LGWKSSYGTGKDAITSGIEVVTNTPTKWDNSFLEILYGYEWELTKSPAGAWYTKAD 357

362 GAGAGTIPDAH-DPSKSHAPFMTTDLALRMDPDYEKISRYYENPDFAFAKAWYKL 420

358 GAGAGTIPDPFGGPGRS--PTMLATDLSLRVDPIYERITRWLEHPEELADEFAKAWYKL 415

421 THRDMPKRVYLGPEVPOEDLIWQDPIPDVSHPLVDENDIEGLKAKILESGLTVSELVST 480

416 THRDMPGVARYLGPLVPKQTLWQDPVPAVSHDLVGEAEIASLSQIRASGLTVSQLVST 475

481 AWASASTFRNSDKRGANGARIRLAPQKDWENNPQ-OLARVLTKEGIEQEDFNOAQSDN 539

476 AWAAASSFRGSDKRGANGRRIRLQPVQWENNDPDGDLRKVIRTLERIEQSFNSAAPGN 535

540 KAVSLADILVLGACAGVEKAADAGHEVOVPFNPGRADATAPQTDVEAFEALEPAADGFR 599

536 IKVSFADLVVLGGCAAIEKAAKAAGHNTVPTPGRTDASQEQTDVESFAVLEPKADGFR 595

600 NYIKPEHKVSAEEMLVDRALQLLSAPENTALVGGMRVLTGNYDGSQHGVTNKGQOLSN 659

596 NYLKGKGNLPAPYMLDKANLLTSLAPENTVLVGGRLVGLGANYKRLPLGVFTFSESELTN 655

660 DFFVNLDLNTKWRASDESDEKVFEGRDKTKTEGVKWSGTRVDLIFGNSSELRALAVYGA 719

656 DFFVNLDMGWTWESPADDTYQKGD-GSGKVKWTGSRVDLVFGNSSELRALVEYVGD 714

720 DSEKFKVDKVKAAKVMMDLDRFDLK 745

715 DAQPRFVQDFVAAMDKVMNLDREFVR 740

RESULT 3

AAW31343

ID AAW31343 standard; Protein; 740 AA.

AC AAW31343;

XX AAW31343;

XX 20-MAR-1998 (first entry)

XX XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

DE XX

QY 660 DFFVNLDDLTNKKWRASDESKVPEGRDFKTEVKNWSTGRVLDLFGSNSELSALAEVYGC 719
 Db 656 DFFVNLDDMGITWEPSPADDTGYGKD-GSGKVKWTGSRVDLVFGSNSELSALAEVYGC 714
 QY 720 DSEKFKVDFKAWAKVMDLDRFDLK 745
 Db 715 DAQPKFVQDFVAWVKVMDLDRFDVR 740

RESULT 4

AAW95398

ID AAW95398 standard; Protein; 740 AA.

XX AAW95398;

XX 26-MAR-1999 (first entry)

XX M. tuberculosis catalase peroxide (katG) gene product.

XX Catalase-peroxide; katG; mutant; mutation; isonicotinic acid hydrazide;

XX INH; tuberculosis; diagnosis; detection.

XX Mycobacterium tuberculosis.

XX W09850585-A1.

XX 12-NOV-1998.

XX 06-MAY-1998; 98WO-US09285.

XX 07-MAY-1997; 97US-0852219.

XX (MAYO-) MAYO FOUNDATION.

XX Cockerill FR, Kline BC, Uhl JR;

XX WPI; 1998-070099/06.

XX N-PSDB; AAX00817.

XX Detection of Mycobacterium tuberculosis - by amplifying katG gene

XX and detecting specific fragment, and optionally identifying

XX INH-resistant strains by detecting specific mutation

XX Claim 3; Fig 7; 83pp; English.

XX The invention relates to a novel method of detecting Mycobacterium

XX tuberculosis. The method comprises amplifying the DNA in the samples to

XX generate a detectable amount of amplified DNA comprising a catalase-

XX peroxide (katG) DNA fragment with sequence of bases 904-1523 of the

XX M. tuberculosis strain H37Rv MC katG gene sequence. The method optionally

XX further comprises determining if the katG DNA fragment has a serine to

XX threonine mutation in codon 315 (S315T mutation), indicative of an

XX isonicotinic acid hydrazide (INH)-resistant M. tuberculosis strain. The

XX method can be used to detect M. tuberculosis in biological fluids,

XX especially human sputum, useful to diagnose tuberculosis. This disease is

XX a major cause of human morbidity and mortality, and conclusive diagnosis

XX and subsequent treatment depends on identification of the etiologic agent

XX M. tuberculosis. INH has been used in tuberculosis treatment, but

XX INH-resistant strains have emerged; the method allows such drug-resistant

XX strains to be identified. The present sequence represents the wild-type

XX M. tuberculosis katG gene product.

XX Sequence 740 AA;

SQ

Query Match 63.6%; Score 2544; DB 20; Length 740;

Best Local Similarity 64.3%; Pred. No. 1.4e-215;

Matches 480; Conservative 88; Mismatches 168; Indels 10; Gaps 8;

QY 2 ENHKGSSSTYNTGCKPFTGGSLKQSGAGTKNRDMPNMLNGILRQHSSLSDPND 61

Db 3 EQHPITETTTGAASNG-CPVVGHMKYPEVGG--NQDWWPNRLNLKVLHQNPAVADPMG 59

QY 62 PDFDYAEFKKLDLAALKVMDLDRFDVR 740

Db 60 AAFDYAAEATIDVDALTRDIEEVMTTSQWPNADYGHYGLFIRMAWHAAGTYRIHDGR 119
 QY 122 GGGSGSQRPAPLNSWPDNANLDKRLLPKOKYGRKISWADLMLLTGNVALETWGF 181
 Db 120 GGAGGGQRPAPLNSWPDNANLDKRLLPKOKYGRKISWADLMLLTGNVALETWGF 179
 QY 182 TFGFAGGRADVWPEEDVYWGAEETEWIGDKRYEGDRELENPLAGAVQMGILYVNPGEANGK 241
 Db 180 TFGFAGGRADVWPEEDVYWGAEETEWIGDKRYEGDRELENPLAGAVQMGILYVNPGEANGK 238
 QY 242 PDFIAAARDIRETFGRMAMNDEETVALIAGHGTGKTHGAADAKEYVGRPAAGIEEMS 301
 Db 239 PDFMAAAVDIRETFGRMAMNDEETVALIAGHGTGKTHGAADAKEYVGRPAAGIEEMS 297
 QY 302 LGHKNTYGTGHGADTITSGLEGAWTKTPTOWSNFFENLFGYEWELTKSPAGAYQWPKD 361
 Db 298 LGHKNTYGTGHGADTITSGLEGAWTKTPTOWSNFFENLFGYEWELTKSPAGAYQWPKD 357
 QY 362 GAGAGTIPDAH-DPSKSHAPFMTLTDLALRMDPDYKISRYYENPDFAFAKAWYKL 420
 Db 358 GAGAGTIPDPFGGPGRS--PTMLATDLSLRVDPYERITRRWLEHPEELADEFAKAWYKL 415
 QY 421 THRMGPKVRYLGPVPEQDLINODPIPDVSHPLVDENDIEGLKAKILESGLTVSELVST 480
 Db 416 IHRDMGPFVARYLGPLVPKQTLWQDPVPVAVSHDLVGEAEIASLKQTRASGLTVSGLVST 475
 QY 481 AWASASTFRNSDKRGANGARIRLAPQKDWENNPQ-QLARVLKLTGEGIOEDFNQASDN 539
 Db 476 AWASASTFRNSDKRGANGARIRLAPQKDWENNPQ-QLARVLKLTGEGIOEDFNQASDN 535
 QY 540 KAYSLADLIYLAGCAGVEKAAKDAGHEVQVFPNPGRADATAEQTDVEAFPALEPAADGFR 599
 Db 536 IKVSFADLVYLGCAAEKAAKAAGHNITVPTFGRTDASQEQTDVESFAVLEPKADGFR 595
 QY 600 NYIKPEHKVSAEMLVDRAOLLSLSAPENTALVGGMRVLGTNTYDGGHGVFTTKPKQOLSN 659
 Db 596 NYLGKGNPLPAEYMLLDKANLLTSAPEMTVLVGLRVLGANYKRLPLGVFTSEASBLTN 655
 QY 660 DFFVNLDDLTNKKWRASDESKVPEGRDFKTEVKNWSTGRVLDLFGSNSELSALAEVYGC 719
 Db 656 DFFVNLDDMGITWEPSPADDTGYGKD-GSGKVKWTGSRVDLVFGSNSELSALAEVYGC 714
 QY 720 DSEKFKVDFKAWAKVMDLDRFDLK 745
 Db 715 DAQPKFVQDFVAWVKVMDLDRFDVR 740

RESULT 5

AAW95399

ID AAW95399 standard; Protein; 740 AA.

XX AAW95399;

XX 26-MAR-1999 (first entry)

XX M. tuberculosis catalase peroxide (katG) gene variant.

XX Catalase-peroxide; katG; mutation; isonicotinic acid hydrazide;

XX INH; tuberculosis; diagnosis; detection; variant.

XX Mycobacterium tuberculosis.

XX Key Location/Qualifiers

XX Misc-difference 315

XX /label= S315T

XX /note= "wild-type Ser is replaced by Thr"

XX W09850585-A1.

XX 12-NOV-1998.

XX 06-MAY-1998; 98WO-US09285.

[illegible]

QY	481	AWASASTFRNSDKRGANGARIRLAPQKDMVNNPQ-QLARVLKLTLEGIQEDFNQAQSDN	538
Db	476	AWAAASFRGSDKRGANGRRIRLQPOVGMEVNDPDGDLRKVIRTLLEEIQESNSAAPGN	535
QY	540	KAVSLADILIVLAGCAGVEKAKDAGHEVOVPFENPCRADATAEQTDYEAPEALEPADGFR	599
Db	536	IKVSPADLIVLGGCAAIKAAKAAGHNITVFETPGRTDASOBQTDVESFAVLEPKADGFR	595
QY	600	NTIKPEKHYSAEMLVDRAQLLSAPEMTALVGGMRVLGTYNDGSHGVFTNKPQOLS	659
Db	596	NYLKGKNPLPAEYMLLDKANLLTISAPEMTVLVGGLAVLGANKRPLGVFTASESLTN	655
QY	660	DFEVNLLDLNTKWRASDESQKVFEGRDFTKTEVSKWSTGYDLIFGNSSELRALAEYVGA	719
Db	656	DFEVNLLDMGITWEPSPADDTGYQCKD-GSGKVKWTGSRVDLVFGNSSELRALVEVYGAD	714
QY	720	DSEEEKFVKDFYKAWAKVMDLDRFDLK	745
Db	715	DAQPKFVQDFAAWDKVNNLDRFDVR	740
RESULT	6		
AAW33809			
ID	AAW33809	standard: Protein; 753 AA.	
XX	AAW33809;		
AC	AC		
DT	DT	18-NOV-1998 (first entry)	
XX	XX		
DE	DE	Alcaligenes (Deleya) aquamarinus catalase-64CA2.	
XX	XX	Catalase; epoxidation; hydroxylation; biosensor; paper bleaching;	
KW	KW	pasteurisation.	
XX	XX		
OS	OS	Alcaligenes aquamarinus.	
XX	XX		
PN	PN	WO9800526-A1.	
XX	XX		
PD	PD	08-JAN-1998.	
XX	XX		
PF	PF	03-JUL-1997; 97WO-US16513.	
XX	XX		
PR	PR	03-JUL-1996; 96US-0674887.	
XX	XX		
PA	PA	(RECO-) RECOMBINANT BIOCATALYSIS INC.	
PI	PI	Adhikary RS, Robertson DE, Sanyal I;	
XX	XX		
DR	DR	WPI; 1998-086953/08.	
XX	XX	N-PSDB; AAV06554.	
XX	XX		
PT	PT	New bacterial catalases, related nucleic acid vectors and	
PT	PT	transformed cells - used as oxidising agents and for detecting or	
XX	XX	destroying hydrogen peroxide, e.g. in biosensors	
PS	PS	Claim 1; Fig 1; 35pp; English.	
XX	XX		
CC	CC	The present sequence is of the Alcaligenes aquamarinus catalase-64CA2.	
CC	CC	Catalase-64CA2 may be used to catalyse oxidation reactions such as	
CC	CC	epoxidation or hydroxylation. The enzyme can also be used to detect or	
CC	CC	destroy hydrogen peroxide, e.g. in connection with glyoxylic acid	
CC	CC	production, biosensors, contact lens cleaning, pulp/paper bleaching and	
CC	CC	pasteurisation of dairy products. Antibodies raised against	
CC	CC	catalase-64CA2 can be used to screen libraries for detection and	
CC	CC	purification of cells containing the enzyme. Fragments of the gene	
XX	XX	encoding Catalase-64CA2 can be used to identify related sequences.	
SQ	SQ	Sequence 753 AA;	

QY 18 GKCPETGSLKOSAGGGTKNRDWPMLNLGLRQHSLSLSDPNDPDDYAEFEKLDLAA 77
 Db 38 GKCPVHGG---NTSTGTSNKDWPPEGLNLDLHQDRKSDPMDPDFNYREEVRKLDLDA 94
 QY 78 VKKDLAALMTSDQDWWPADIYGHYGFPIRMWHSAGTYRIGDGRGGSGSQRAPLNSW 137
 Db 95 LKQDVHALMTDSQEWMPADWGHYGLLMIRMAHWSAGTYRIADGRGGGTGSGRQAPLNSW 154
 QY 138 PDNANLAKARLLLPKIKYGRKTSWADMLTLGNVALETMGFTKFGAGGRADWPEPE 197
 Db 155 PDNVSLDKARRLLWPDKKYGKISWADMLTLAGTVAYESMGLPAYGFSFGVRDIWEPEK 214
 QY 198 DVTWGAETEWL--GDKRYEGD----RELENPLGAVOMGLIYVNPENPGKDPDPAARDI 251
 Db 215 DIYWGDEKELAPSDERY-GDVNKPETMENPLAAVOMGLIYVNPENPGKDPDPLRTAQOV 273
 QY 252 RETFGRMAMDEETVALIAGHTFGKTHGAADAQKYVGRPEAAAGIEMSLGWKNYCTG 311
 Db 274 LETFARMAMDEKTAALTAGGTVGNCHGNASA-LAPDPKASDVENOGGLWGNPNMQG 332
 QY 312 HGADTITSGLEGAWTKTPTQNSNFFENLFGYEWELTKSPAGAYQWKPKDGAGAGTIPDA 371
 Db 333 KASNAVTSIGIEGAWTNTKFDMGYFDLLFGYNWELKSPAGAHWEPIIDIKKENKPVDA 392
 QY 372 HDPSKSHAPFMTTDLALRMDPDYEKISRYYENPDEFAFAKAWYKLTDRMGPKVRY 431
 Db 393 SDPSIRHNPIMTDADMAIKVNPTYRAICEKFMADPEYFKTKFAKAWFKLTDRDLGPKSRY 452
 QY 432 LGPEVPQBDLIQWDPIDPVSHPLVDENDIEGLKAKILESGLTVSELVSTANASASTPNS 491
 Db 453 IGPEVPAEDLLIQWDPIDAGNTDYCE---EVVKQKIASGSLISEKSVSTANDSARTYRGS 508
 QY 492 DKRGANGARILAPQKDWYNNPQOLARVLKTLLEGIOEDFNQAQSDNKAYSVLADLIVLA 551
 Db 509 DMRGANGARILAPQNEWQNEPERLAKVLSVEQISAD-----TGASIAADVILA 560
 QY 552 GCAGVEKAAKAGHEVQVPPNPGRADATAEOTDVEAFEALEPAADGFERNYIKPEKVSAAE 611
 Db 561 GSVGIEKAAKAGDVRVFFKUGKGDATAEOTDVEAFEALEPAADGFERNYIKPEKVSAAE 620
 QY 612 EMLVDRAQLLSLSAPEMTALYGGMRVLGTNTYDGSQHGVTNKPQQLSNDFFVNLJDLNFK 671
 Db 621 EMLLDRAQLMGLTGPMTVLGMRVLGTNTYDGTGKHGVTETDCEGLTNDFFVNLTDGNS 680
 QY 672 WRASDESKVEGRDFTGEVKKWGTNRVLDLFGNSSELRAEAIVYGCADSEKFKVDFVK 731
 Db 681 WK--PVGSNAYEIRDKRTGAVKWTASRYDLVFGSNLSLRSYAEVYQAQDNGEKEFVRDVA 738
 QY 732 AWAKVMDLDRFDL 744
 Db 739 AWTVMNADRFV 751
 RESULT 7
 ABG33040
 ID ABG33040 standard; Protein; 753 AA.
 XX AC ABG33040;
 XX 22-NOV-2002 (first entry)
 XX DE Alcalligenes catalase protein 64CA2.
 XX Gene; ss; marine bacteria; catalase; hydrogen peroxide; enzyme.
 XX OS Alcalligenes aquamarinus.
 XX PN US2002102680-A1.
 XX PD 01-AUG-2002.
 XX PF 19-JUN-2001; 2001US-0884889.
 XX

PR 03-JUL-1996; 960S-0674887.
 PR 16-OCT-1997; 970S-0951844.
 PR 05-OCT-1999; 990S-0412347.
 XX (ROBE/) ROBERTSON D E.
 PA (SANT/) SANYAL I.
 PA (ADHI/) ADHIKARI R.
 XX Robertson DE, Sanyal I, Adhikari R;
 PI WPI: 2002-690613/74.
 DR N-PSDB: ABS53944.
 XX
 PT New purified catalase polypeptide useful for catalyzing the breakdown
 of hydrogen peroxide, for modifying small molecules, and for generating
 antibodies which bind to the polypeptide
 PT Claim 22; Plg 5; 44pp; English.
 PS
 XX This invention relates to the DNA and protein sequences of a novel
 purified catalase polypeptide. The protein of the invention or its
 homologue is useful for catalyzing the breakdown of hydrogen peroxide.
 CC The protein sequence of the invention is useful for modifying small
 CC molecules, by mixing the protein or its fragments with a small molecule
 CC to produce a modified small molecule. The protein of the invention is
 CC also useful for catalyzing biochemical reactions, for hydrolyzing
 CC glycosidic linkages and for generating antibodies which bind
 CC specifically to the protein. The nucleic acid sequences of the
 CC invention is useful as a probe to determine whether a biological sample,
 CC such as a soil sample, contains an organism having the nucleic acid or
 CC an organism from which the DNA was obtained, or for identifying an
 CC isolating related nucleic acids. An antibody that binds the cellulase
 CC protein of the invention is useful in immunoaffinity chromatography
 CC procedures to isolate or purify the protein or for detection of protein
 CC expression in a biological sample. The cellulase protein of the
 CC invention is heat stable, is heat resistant, and is able to
 CC renature and regain activity after exposure to temperatures of from
 CC about 60 to 105 degrees. The present sequence represents the Alcalligenes
 CC aquamarinus catalase protein of the invention.
 XX
 SQ Sequence 753 AA;
 Query Match 57.2%; Score 2290.5; DB 23; Length 753;
 Best Local Similarity 59.1%; Pred. No. 3.8e-193;
 Matches 433; Conservative 92; Mismatches 183; Indels 25; Gaps 8;
 QY 18 GKCPETGSLKOSAGGGTKNRDWPMLNLGLRQHSLSLSDPNDPDDYAEFEKLDLAA 77
 Db 38 GKCPVHGG---NTSTGTSNKDWPPEGLNLDLHQDRKSDPMDPDFNYREEVRKLDLDA 94
 QY 78 VKKDLAALMTSDQDWWPADIYGHYGFPIRMWHSAGTYRIGDGRGGSGSQRAPLNSW 137
 Db 95 LKQDVHALMTDSQEWMPADWGHYGLLMIRMAHWSAGTYRIADGRGGGTGSGRQAPLNSW 154
 QY 138 PDNANLAKARLLLPKIKYGRKTSWADMLTLGNVALETMGFTKFGAGGRADWPEPE 197
 Db 155 PDNVSLDKARRLLWPDKKYGKISWADMLTLAGTVAYESMGLPAYGFSFGVRDIWEPEK 214
 QY 198 DVTWGAETEWL--GDKRYEGD----RELENPLGAVOMGLIYVNPENPGKDPDPAARDI 251
 Db 215 DIYWGDEKELAPSDERY-GDVNKPETMENPLAAVOMGLIYVNPENPGKDPDPLRTAQOV 273
 QY 252 RETFGRMAMDEETVALIAGHTFGKTHGAADAQKYVGRPEAAAGIEMSLGWKNYCTG 311
 Db 274 LETFARMAMDEKTAALTAGGTVGNCHGNASA-LAPDPKASDVENOGGLWGNPNMQG 332
 QY 312 HGADTITSGLEGAWTKTPTQNSNFFENLFGYEWELTKSPAGAYQWKPKDGAGAGTIPDA 371
 Db 333 KASNAVTSIGIEGAWTNTKFDMGYFDLLFGYNWELKSPAGAHWEPIIDIKKENKPVDA 392
 QY 372 HDPSKSHAPFMTTDLALRMDPDYEKISRYYENPDEFAFAKAWYKLTDRMGPKVRY 431
 Db 393 SDPSIRHNPIMTDADMAIKVNPTYRAICEKFMADPEYFKTKFAKAWFKLTDRDLGPKSRY 452

QY 525 LEG-----IQDFNOASDNKAVSL-----ADLIVLAGC-----AGVEKAAKDAGHEVQV 569
 Db 511 LRGHPARTAEKSKHRLDRGLTGRKQPAATPALMSNCHFLSAAAMRHKSQPMKALPC 570
 QY 570 PFNPGRADATAEQDVEAFALPAADGPFNRYIKPEKVSAAEMLVDRAQLLSAPEWT 629
 Db 571 WNRQSASAT-----IKSKTRFRKSCSSTKPSS-----SADRPRND 608
 QY 630 ALVGGMRVLGTNYDGHGVTNKGQSLNSDFVNLDDLTNKKRASDESKVFEGRDKFT 689
 Db 609 GLSWRFARVCPNRYHLPHGVFTDRIGVLTNDFFVNLDDMYEWPVDSG--IYEIRDKT 666
 QY 690 GEVWKSSTRVDLIFGNSNELRALAEVYGCADSEKFKVDFVKAWKVMJLDRFDL 744
 Db 667 GEVWTRATRVLDLIFGFYQADDNQEERFYAODDNOEKFVRDPINAVWVKNADREDL 721

RESULT 12
 ABG25057
 ID ABG25057 standard; Protein: 533 AA.
 XX AC ABG25057;
 XX DT 18-FEB-2002 (first entry)
 XX DE Novel human diagnostic protein #25048.
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX OS Homo sapiens.
 XX PN WO200175067-A2.
 XX PD 11-OCT-2001.
 XX PF 30-MAR-2001; 2001WO-US08631.
 XX PR 31-MAR-2000; 2000US-0540217.
 XX PR 23-AUG-2000; 2000US-0649167.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Drmanac RT, Liu C, Tang YT;
 XX DR WPI; 2001-639362/73.
 XX DR N-PSDB; AAS89244.
 XX PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX PS Claim 20; SEQ ID No 55416; 103pp; English.
 XX CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 533 AA;
 QY Query Match 31.6%; Score 1266; DB 22; Length 533;
 Best Local Similarity 60.9%; Pred. No. 9.4e-103;
 Matches 255; Conservative 55; Mismatches 95; Indels 14; Gaps 6;
 QY 328 TPQWSNNFENLFGYEWELTKSPAGAYQWKPKDAGAGTIPDAHDSKSHAPMLTDL 387
 Db 126 TPQWSNYPFENLFGYEWVQTRSPAGAIQFEAVD--APEIIPDPFDSKKRKTMLTDL 183
 QY 388 ALRMDPYEKISRRYENPDEFAFAKANYKLTTHRDMGPKRVYLGPEVQEDLIWDPI 447
 Db 184 TLRFDPEFKISRRLNDPQAFNEAFARWFKLTHRDMGPKRSRYIGPEVKEDLIWDPL 243
 QY 448 PD-VSHPLVDNDIEGLKAKILSEGLTVSELVSTAWASASTFRNSDKRGANGARILAP 506
 Db 244 PQIYNP--TEQDIIDLKFAIDSLVSELVSVAMASASTFRGDKRGANGARLAMP 301
 QY 507 Q-KDWEVNNPQQLARVLKTLLEGIQEDPNQASDNKAVSLADLIVLAGCAGVEKAAKDAGH 565
 Db 302 QTRTGDVTRRQPI-RALPVLKIQKESGKA-----SLADIIVLPGLFVVEKSSAAGL 353
 QY 566 EVQVPFNPGRADATAEQDVEAFALPAADGPFNRYIKPEKVSAAEMLVDRAQLLSA 625
 Db 354 SIHVFPAPGRVARDARQDQTALGMFELLEPIADGPFNRYARLDVSTESLLIDKAOQLTLTA 413
 QY 626 PENTALVGGMRVLGTNYDGHGVTNKGQSLNSDFVNLDDLTNKKRASDESKVFEGR 685
 Db 414 PENTALVGGMRVLGANGFNGSKNGVFTDRVGLNSDFFVNLDDMYEWKATDESKLEFGR 473
 QY 686 DFKTGEVKWSGTRVDLIFGNSNELRALAEVYGCADSEKFKVDFVKAWKVMJLDRFDL 744
 Db 474 DRETGEVKFTASRADLVFGNSVLRVAEYVASSDAHEKFKDFVAAWVKVMJLDRFDL 532

RESULT 13
 ABG08823
 ID ABG08823 standard; Protein: 273 AA.
 XX AC ABG08823;
 XX DT 13-FEB-2002 (first entry)
 XX DE Novel human diagnostic protein #8814.
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX OS Homo sapiens.
 XX PN WO200175067-A2.
 XX PD 11-OCT-2001.
 XX PF 30-MAR-2001; 2001WO-US08631.
 XX PR 31-MAR-2000; 2000US-0540217.
 XX PR 23-AUG-2000; 2000US-0649167.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Drmanac RT, Liu C, Tang YT;
 XX DR WPI; 2001-639362/73.
 XX DR N-PSDB; AAS73010.
 XX PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess
XX biodiversity -
PS Claim 20; SEQ ID No 39182; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 273 AA;
Query Match 22.1%; Score 883.5; DB 22; Length 273;
Best Local Similarity 62.6%; Pred. No. 2.2e-69;
Matches 176; Conservative 39; Mismatches 57; Indels 9; Gaps 2;
QY 464 KAKILESGLTVSELVSTAWASASTFRNSDKRGANGARIRLAPQKDWYVNFQOLARVLK 523
DB 1 KFAIDLSGLSVSELVSVANASASTFRGDKRGANGARLALMPQRDWYN--AAAVRALP 58
QY 524 TLEGTOEDFNQASDNKAVSLADLIVLACGAGVEKAKADAGHEVQVFFNPGRADATAEQT 583
DB 59 VLEKIQESGKA-----SLADIIVLAGVGVGVEKAAAGLSIHVPFAGPRVDARQDQT 111
QY 584 DYAEFALEPAADGFNRYIKPEHKVSAEEMLVDRQALLSLAPENTALVGGMRVLGTYND 643
DB 112 DIEMFELLEPIADGFNRYARLDVSTESLLIDKAQQLITAPENTALVGGMRVLGANTFD 171
QY 644 GSOHGVFTNKPQGLSNDFFVNLNLDLNTKWRASDESKVFEGRDFTKGVKSGTRVDLIF 703
DB 172 GSKNGVFTDRGVLSNDFVNLDMRYENKATDESKELFEGRDRETGENKFTASRADLVE 231
QY 704 GSNSELRALEYVGCADSEKFKVDFKAWAKVMDLDRFDL 744
DB 232 GSNVLRAVLEYVASSDAHEKFKVDFVAAWVKVMNLDLDRFDL 272
RESULT 14
ABG09297
ID ABG09297 standard; Protein; 1550 AA.
XX
AC ABG09297;
XX
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #9288.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
XX
PD 11-OCT-2001.
XX

PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
PI
DR WPI; 2001-639362/73.
DR N-PSDB; AAS73484.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 39656; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1550 AA;
Query Match 16.9%; Score 675; DB 22; Length 1550;
Best Local Similarity 60.9%; Pred. No. 1e-49;
Matches 137; Conservative 24; Mismatches 50; Indels 14; Gaps 5;
QY 328 TPTQWNNFFENLFGYEWELTKSPAGAYOWKPKDGAGAGTIPDAHPDPSKSHAPFMTLTDL 387
DB 528 TPTQWNNFFENLFGYEWELTKSPAGAYOWKPKDGAGAGTIPDAHPDPSKSHAPFMTLTDL 585
QY 388 ALRMDPDYKISRYYENPDEFADAFKAWYKLTDRMGPKVRYLGPVPEQDLIMQDPI 447
DB 586 TLRFDPFEKISRRLNDPQAFNEAFARAFKLTDRMGPKVRYLGPVPEQDLIMQDPI 645
QY 448 PD-VSHPLVDNDIEGLKAKILESGLTVSELVSTAWASASTFRNSDKRGANGARIRLAP 506
DB 646 PQPIYNP--TEQDIIDLKFAIDLSGLSVSELVSVANASASTFRGDKRGANGARLALMP 703
QY 507 QKDWYVNFQOLARVLKLTLEGTOEDFNQASDNKAVSLADLIVLA 551
DB 704 QRDWDVN--AAAVRALPVLKIQESGKA-----SLADIIVLA 739
RESULT 15
ABG08828
ID ABG08828 standard; Protein; 201 AA.
XX
AC ABG08828;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #8819.

```

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX N-PSDB; AAS73015.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX Claim 20; SEQ ID No 39187; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 201 AA;
Query Match 11.4%; Score 455; DB 22; Length 201;
Best Local Similarity 42.5%; Pred. No. 1.1e-31;
Matches 105; Conservative 26; Mismatches 42; Indels 74; Gaps 4;
QY 468 LESGLTVSELYSTAWASASTFRNSDKRGGANGARIRLAPQKDEWVNNPQQLARVLKLTIEG 527
Db 4 LPSLFTFTRLSS-----SDNDLRGGDKRGGANGARLALXPKQ-----TFTG 43
QY 528 IQEDFNQAQDNKAVSLADLIVLAGCAGVEKAADAGHEVOVFPNPGRADATAEQ----- 582
Db 44 -----DVTTRQPIRVC 54
QY 583 TDVAFALPAPADGFNRYIKPEHKVSAEMLVDRAQLLSLAPENTALVGGMRVLGTNY 642
Db 55 TVIEMFELLEPIADGFNRYARLDVSTESLLDKAQLTLTAPENTALVGGMRVLGANF 114
QY 643 DGSQHGVTNKPGLSNDFFVNLLDLNTKWRASDESKVFEGRDFTKEVKNWCGTRVDLI 702
Db 115 DGSKNGVFTDRVGLSNDFFVNLLDMRYENKATDESKELFEGRDRETGEVKNFTASRADLV 174

```

QY 703 FGSNSEL 709
 Db 175 FGSNSVL 181

Search completed: October 7, 2003, 19:42:23
 Job time : 63.1769 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 7, 2003, 19:40:18 ; Search time 21.8825 Seconds
(without alignments)
1440.492 Million cell updates/sec

Title: US-09-884-889-8
Perfect score: 4002
Sequence: 1 MENHKSSSYNTTGGKC.....VKDFKAKAKYMDLRDLK 745

Scoring table:
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA:
1: /cgn2_6/ptodata/1/laa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/laa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/laa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/laa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/laa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/laa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4002	100.0	745	2	US-08-674-887A-8
2	4002	100.0	745	3	US-08-951-844-8
3	4002	100.0	745	4	US-09-412-347-8
4	2344	63.6	740	1	US-08-418-782-7
5	2344	63.6	740	1	US-08-418-782-21
6	2344	63.6	740	1	US-08-228-662-7
7	2344	63.6	740	2	US-08-852-219-7
8	2344	63.6	740	2	US-08-852-219-21
9	2421.5	60.5	726	2	US-08-313-185-49
10	2421.5	60.5	726	2	US-08-459-499-13
11	2421.5	60.5	726	3	US-09-082-614A-49
12	2397	59.9	729	2	US-08-313-185-50
13	2397	59.9	729	3	US-08-459-499-14
14	2397	59.9	729	3	US-09-082-614A-50
15	2331	58.2	726	4	US-09-328-352-8115
16	2290.5	57.2	753	2	US-08-674-887A-6
17	2290.5	57.2	753	3	US-08-951-844-6
18	2290.5	57.2	753	4	US-09-412-347-6
19	2278.5	56.9	735	2	US-08-313-185-48
20	2278.5	56.9	735	3	US-08-459-499-9
21	2278.5	56.9	735	3	US-09-082-614A-48
22	2270.5	58.7	735	2	US-08-459-499-12
23	1954	48.8	652	2	US-08-313-185-53
24	1954	48.8	652	2	US-08-459-499-17
25	1954	48.8	652	3	US-09-082-614A-53
26	1760.5	44.0	731	2	US-08-313-185-51
27	1760.5	44.0	731	2	US-08-459-499-15

28	1760.5	44.0	731	3	US-09-082-614A-51
29	312.5	7.8	293	2	US-08-313-185-52
30	312.5	7.8	293	3	US-09-082-614A-52
31	311	7.8	294	2	US-08-459-499-16
32	219	5.5	78	1	US-07-929-206-4
33	219	5.5	78	2	US-08-313-185-44
34	219	5.5	78	2	US-08-459-499-4
35	219	5.5	78	3	US-09-082-614A-44
36	179	4.5	37	1	US-07-929-206-6
37	179	4.5	37	2	US-08-313-185-3
38	179	4.5	37	2	US-08-459-499-6
39	179	4.5	37	3	US-09-082-614A-3
40	172	4.3	37	1	US-07-929-206-5
41	172	4.3	37	1	US-07-929-206-8
42	172	4.3	37	2	US-08-313-185-2
43	172	4.3	37	2	US-08-459-499-5
44	172	4.3	37	3	US-09-082-614A-2
45	138.5	3.5	120	1	US-07-929-206-2

ALIGNMENTS

RESULT 1
US-08-674-887A-8
; Sequence 8, Application US/08674887A
; Patent No. 5939300
; GENERAL INFORMATION:
; APPLICANT: Robertson, Dan E.
; APPLICANT: Sanval, Indrajit
; APPLICANT: Adhikari, Robert S.
; TITLE OF INVENTION: CATALASES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/674,887A
; FILING DATE: 03-JUL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09015/002001
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 745 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: Internal
US-08-674-887A-8

Query Match 100.0%; Score 4002; DB 2; Length 745;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 745; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 MENHKSSSYNTTGGKCPTGSLKOSAGGTTKRDWPNMLNLGLIRHSSLSDPN 60
|||||

```
Db 1 MENKHSGSSTYNTNTGGCKPFTGGSLKQSGAGGTTKRDWPNMNLGILRQHSLSLDPN 60
QY 61 DPFDYAEFPKLLDLAAVKKDLAALMTDSQDWNPADYGHYGPFTIRMAHWSAGTYRIGDG 120
Db 61 DPFDYAEFPKLLDLAAVKKDLAALMTDSQDWNPADYGHYGPFTIRMAHWSAGTYRIGDG 120
QY 121 RGGGSGSORFAPLNSWPDNANLDKARLLWPIKQYGRKISWADLMILTNVALETWGMF 180
Db 121 RGGGSGSORFAPLNSWPDNANLDKARLLWPIKQYGRKISWADLMILTNVALETWGMF 180
QY 121 RGGGSGSORFAPLNSWPDNANLDKARLLWPIKQYGRKISWADLMILTNVALETWGMF 180
Db 121 RGGGSGSORFAPLNSWPDNANLDKARLLWPIKQYGRKISWADLMILTNVALETWGMF 180
QY 181 KTFGAGGRADVWPEEDVYWGAEETWLGDKRYEGDRELENPLGAVOMGLIYVNPPEGNG 240
Db 181 KTFGAGGRADVWPEEDVYWGAEETWLGDKRYEGDRELENPLGAVOMGLIYVNPPEGNG 240
QY 241 KPDIAAARDIRETFGRMANDEETVALIAGGHTFGKTHGAADAEEKYVGREPAAAGTEEM 300
Db 241 KPDIAAARDIRETFGRMANDEETVALIAGGHTFGKTHGAADAEEKYVGREPAAAGTEEM 300
QY 301 SLGKNTYGTGHGADTTTSGLEGAWTKPTQWSNFFENLFGYEWELTKSPAGAYQWPK 360
Db 301 SLGKNTYGTGHGADTTTSGLEGAWTKPTQWSNFFENLFGYEWELTKSPAGAYQWPK 360
QY 361 DGAGAGTIPDAHDPKSHAPFMTLTDLALRMDPDYKISRYYENPDEFADAFKAWYKL 420
Db 361 DGAGAGTIPDAHDPKSHAPFMTLTDLALRMDPDYKISRYYENPDEFADAFKAWYKL 420
QY 421 THRDMPKRYLGPVEPQEDLIWQDPIPDVSHPLVDNDIEGLKAKILESGLTVSELVST 480
Db 421 THRDMPKRYLGPVEPQEDLIWQDPIPDVSHPLVDNDIEGLKAKILESGLTVSELVST 480
QY 481 AWASASTFRNSDKRGANGARIRLAPOKDWEVNNPQOLARVLKTEGIDEDFNOAQSDNK 540
Db 481 AWASASTFRNSDKRGANGARIRLAPOKDWEVNNPQOLARVLKTEGIDEDFNOAQSDNK 540
QY 541 AVSLADLIVLAGCAGVEKAADAGHEVQVFPNPGRADATAEQTDVEAFEALEPAADGFRN 600
Db 541 AVSLADLIVLAGCAGVEKAADAGHEVQVFPNPGRADATAEQTDVEAFEALEPAADGFRN 600
QY 601 YIRPEHKVSAEMLVDRAQLLSAPEMTALVGGMRVLTGNYDGSQHGVTNKPQLSND 660
Db 601 YIRPEHKVSAEMLVDRAQLLSAPEMTALVGGMRVLTGNYDGSQHGVTNKPQLSND 660
QY 661 FFVNLDLNTKWRASDESCKVFEGRDPKTEGVKWSGTRVDLIFGNSSELRALAEVYGCAD 720
Db 661 FFVNLDLNTKWRASDESCKVFEGRDPKTEGVKWSGTRVDLIFGNSSELRALAEVYGCAD 720
QY 721 SEEKFKVDFYKAWKVMDLDRFDLK 745
Db 721 SEEKFKVDFYKAWKVMDLDRFDLK 745
```

RESULT 2

```
US-08-951-844-8
; Sequence 8, Application US/08951844
; Patent No. 6074860
; GENERAL INFORMATION:
; APPLICANT: Robertson et al.
; TITLE OF INVENTION: Catalases
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/951.844
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/674,887
; FILING DATE: July 3, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles J. Herron
; REGISTRATION NUMBER: 28,019
; REFERENCE/DOCKET NUMBER: 331400-55
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 745 AMINO ACIDS
; TYPE: AMINO ACID
; TOPOLOGY: LINEAR
; MOLECULE TYPE:
; US-08-951-844-8
```

```
Query Match 100.0%; Score 4002; DB 3; Length 745;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 745; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MENKHSGSSTYNTNTGGCKPFTGGSLKQSGAGGTTKRDWPNMNLGILRQHSLSLDPN 60
Db 1 MENKHSGSSTYNTNTGGCKPFTGGSLKQSGAGGTTKRDWPNMNLGILRQHSLSLDPN 60
QY 61 DPFDYAEFPKLLDLAAVKKDLAALMTDSQDWNPADYGHYGPFTIRMAHWSAGTYRIGDG 120
Db 61 DPFDYAEFPKLLDLAAVKKDLAALMTDSQDWNPADYGHYGPFTIRMAHWSAGTYRIGDG 120
QY 121 RGGGSGSORFAPLNSWPDNANLDKARLLWPIKQYGRKISWADLMILTNVALETWGMF 180
Db 121 RGGGSGSORFAPLNSWPDNANLDKARLLWPIKQYGRKISWADLMILTNVALETWGMF 180
QY 181 KTFGAGGRADVWPEEDVYWGAEETWLGDKRYEGDRELENPLGAVOMGLIYVNPPEGNG 240
Db 181 KTFGAGGRADVWPEEDVYWGAEETWLGDKRYEGDRELENPLGAVOMGLIYVNPPEGNG 240
QY 241 KPDIAAARDIRETFGRMANDEETVALIAGGHTFGKTHGAADAEEKYVGREPAAAGTEEM 300
Db 241 KPDIAAARDIRETFGRMANDEETVALIAGGHTFGKTHGAADAEEKYVGREPAAAGTEEM 300
QY 301 SLGKNTYGTGHGADTTTSGLEGAWTKPTQWSNFFENLFGYEWELTKSPAGAYQWPK 360
Db 301 SLGKNTYGTGHGADTTTSGLEGAWTKPTQWSNFFENLFGYEWELTKSPAGAYQWPK 360
QY 361 DGAGAGTIPDAHDPKSHAPFMTLTDLALRMDPDYKISRYYENPDEFADAFKAWYKL 420
Db 361 DGAGAGTIPDAHDPKSHAPFMTLTDLALRMDPDYKISRYYENPDEFADAFKAWYKL 420
QY 421 THRDMPKRYLGPVEPQEDLIWQDPIPDVSHPLVDNDIEGLKAKILESGLTVSELVST 480
Db 421 THRDMPKRYLGPVEPQEDLIWQDPIPDVSHPLVDNDIEGLKAKILESGLTVSELVST 480
QY 481 AWASASTFRNSDKRGANGARIRLAPOKDWEVNNPQOLARVLKTEGIDEDFNOAQSDNK 540
Db 481 AWASASTFRNSDKRGANGARIRLAPOKDWEVNNPQOLARVLKTEGIDEDFNOAQSDNK 540
QY 541 AVSLADLIVLAGCAGVEKAADAGHEVQVFPNPGRADATAEQTDVEAFEALEPAADGFRN 600
Db 541 AVSLADLIVLAGCAGVEKAADAGHEVQVFPNPGRADATAEQTDVEAFEALEPAADGFRN 600
QY 601 YIRPEHKVSAEMLVDRAQLLSAPEMTALVGGMRVLTGNYDGSQHGVTNKPQLSND 660
Db 601 YIRPEHKVSAEMLVDRAQLLSAPEMTALVGGMRVLTGNYDGSQHGVTNKPQLSND 660
QY 661 FFVNLDLNTKWRASDESCKVFEGRDPKTEGVKWSGTRVDLIFGNSSELRALAEVYGCAD 720
Db 661 FFVNLDLNTKWRASDESCKVFEGRDPKTEGVKWSGTRVDLIFGNSSELRALAEVYGCAD 720
```

```
QY 721 SEEKFKDFVKAWKVMMDLDRDLK 745
Db 721 SEEKFKDFVKAWKVMMDLDRDLK 745

RESULT 3
; Sequence 8, Application US/09412347
; Patent No. 6410290
; GENERAL INFORMATION:
; APPLICANT: Robertson, Dan E.
; APPLICANT: Sanyal, Indrajit
; APPLICANT: Adhikari, Robert S.
; TITLE OF INVENTION: CATALASES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/412,347
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/674,887
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09015/002001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 745 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: Internal
US-09-412-347-8

Query Match 100.0%; Score 4002; DB 4; Length 745;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 745; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MENHKHSGSSTNTNTGKPTGGSLKOSAGGCTKNRDWPNMLNLGTLRHSSLSDPN 60
Db 1 MENHKHSGSSTNTNTGKPTGGSLKOSAGGCTKNRDWPNMLNLGTLRHSSLSDPN 60

QY 61 DPDFDYAEFFKLDLAANKDLAALMTDSQDWWPADYGHYPFFIRMAWHSAGTYRIGDG 120
Db 61 DPDFDYAEFFKLDLAANKDLAALMTDSQDWWPADYGHYPFFIRMAWHSAGTYRIGDG 120

QY 121 RGGGSGSQRFPAPLNSWPNANLDKARLLLPKQYGRKISWADLMILTNVALETMGF 180
Db 121 RGGGSGSQRFPAPLNSWPNANLDKARLLLPKQYGRKISWADLMILTNVALETMGF 180

QY 181 KTFGAGGADVWPEEDYVWGAETWLGDKRYEGDRELENPLGAVQMGLIYVNPNGP 240
Db 181 KTFGAGGADVWPEEDYVWGAETWLGDKRYEGDRELENPLGAVQMGLIYVNPNGP 240

QY 241 KPDPIAAARDIRETFGRMANDEETVALLAGGHTFGKTHGAADAKEYVVGREPAAGIE 300
Db 241 KPDPIAAARDIRETFGRMANDEETVALLAGGHTFGKTHGAADAKEYVVGREPAAGIE 300

QY 301 SLGWKNTYCTGHGADTITSLGEGAWTKTPTQSNNNFFENLFGYEWELTKSPAGAYQWKPK 360
Db 301 SLGWKNTYCTGHGADTITSLGEGAWTKTPTQSNNNFFENLFGYEWELTKSPAGAYQWKPK 360

QY 361 DGAGAGTIPDAHDPKSHAPFMTTDLALRMDPDYKISRRIYENDEFAFAKAWYKL 420
Db 361 DGAGAGTIPDAHDPKSHAPFMTTDLALRMDPDYKISRRIYENDEFAFAKAWYKL 420

QY 421 THRMGPKVRYLGPVPEQEDLIWQDPIDVSHPLVDENDIEGLKAKILESGLTVSELVST 480
Db 421 THRMGPKVRYLGPVPEQEDLIWQDPIDVSHPLVDENDIEGLKAKILESGLTVSELVST 480

QY 481 AWASASTFRNSDKRGANGARIRLAPQKDWENNNPOOLARVLKTLGIEDFNQAOSDNK 540
Db 481 AWASASTFRNSDKRGANGARIRLAPQKDWENNNPOOLARVLKTLGIEDFNQAOSDNK 540

QY 541 AVSLADLIVLAGCAGVEKAAKADAGHEVQVFPNPGRADATAEQTDFEAFEALEPAADGFRN 600
Db 541 AVSLADLIVLAGCAGVEKAAKADAGHEVQVFPNPGRADATAEQTDFEAFEALEPAADGFRN 600

QY 601 YIKPEKVSAAEMLVDRAQLLSLAPMTALVGGMRVLGTNTDGSQSGHGVFTNNKPGOLSND 660
Db 601 YIKPEKVSAAEMLVDRAQLLSLAPMTALVGGMRVLGTNTDGSQSGHGVFTNNKPGOLSND 660

QY 661 FFVNLDDLTKWRASDESQKVEGRDPKTCGEVWWSGTRVDLIPGWSNSELRAEAETVGCAD 720
Db 661 FFVNLDDLTKWRASDESQKVEGRDPKTCGEVWWSGTRVDLIPGWSNSELRAEAETVGCAD 720

QY 721 SEEKFKDFVKAWKVMMDLDRDLK 745
Db 721 SEEKFKDFVKAWKVMMDLDRDLK 745

RESULT 4
US-08-418-782-7
; Sequence 7, Application US/08418782
; Patent No. 5638733
; GENERAL INFORMATION:
; APPLICANT: Cockerill, Franklin R.
; APPLICANT: Kline, Bruce C.
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Isoniazid Resistant Strains
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg & Woessner
; STREET: 3500 IDS Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,782
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D.
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 150.141US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 740 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

```

; MOLECULE TYPE: protein
US-08-418-782-7

Query Match      63.6%; Score 2544; DB 1; Length 740;
Best Local Similarity 64.3%; Pred. No. 9.5e-222;
Matches 480; Conservative 88; Mismatches 168; Indels 10; Gaps 8;

QY 2 ENHKHSSSYNTNTGKCPPTGSLKQSGAGGKTRNDWPNMLNLGTLQHSLSLSDPND 61
Db 3 EGHPIETTTGAASNG-CPVVGHHKYPVEGG--NQDMPNRLNLKVLHONPAVADPMG 59
QY 62 PDFDYAEFFKDLAAVKDLAALTDSDQWMPADYGHYGFPTFMAHWSAGTYRIGDGR 121
Db 60 AAFDYAAEVATIDVDAITRDIEEVYTTSQPWPADYGHYGLFTRMAHAAAGTYRIHGR 119
QY 122 GGGGSGSORFAPLNSWPNANLDRKLLWPVKYKGRKISWADMLTGNVALETMGFK 181
Db 120 GGAGGGMORFAPLNSWPNASLDKARLLMPVKKYKGLSWADLIYFAGNCALESNGFK 179
QY 182 TFGFAGGRADWPEEDVYGAETWLGDKRYEGDRELENPLGAVQMLIYVNPGEPMGK 241
Db 180 TFGFGRVQWPEDE-VYWGKEATWLGDERYSKRDLENPLAAVQMLIYVNPGEPMGN 238
QY 242 PDPTAAARDIRETFRMAMNDEETVALIAGHTFGKTHGAADAKEYYVGREPAAAGIEMS 301
Db 239 PDPMAAAVDIRETFRMAMNDVETAAALIVGGHTFGKTHGAGPAD-LVGPEPEAAPLEQMG 297
QY 302 LGWKNTYGTGHGADTITSGLEGATWKTPTOWSNFFENLFGYEWELTKSPAGAYQWPKD 361
Db 298 LGWKSSYGTGKDAITSGIEEVVYNTPTKWDNSFLIYGYEWELTKSPAGAWQYTKD 357
QY 362 GAGAGTTPDAH-DPSKSHAPFMTLTDLALRMDPDYKISRYYENPDEFAFAKAWYKL 420
Db 358 GAGAGTTPDPFGGGRS--PTMLATDLSLRVDPYIERITRWLEHPELDEFKAWYKL 415
QY 421 THRDMPKRYLGPVEPQEDLIWQDPIPVSHPLVDENDIEGLKAKILESGLTVELYST 480
Db 416 IHRDMGPVARYLGPLVPKQTLWQDPVAVSHDLVGEAETASLSKQTRASGLTVSQVYST 475
QY 481 AWASASTFRNSDKRGANGARIRLAPQKDWENNPQ-QLARVLKTLLEGIQEDFNQAOSDN 539
Db 476 AWAAASSFRGSKRGANGGRIQLPQVQWGEVNDPDGDLRKVIRTLKEIOESFNSAAGN 535
QY 540 KAVSLADLIYLACAGVEKAKADAGHEVQVFPNPGRADATAEQTDVFAFPALEPAADGFR 599
Db 536 IKVSFADLVVLGGCAAEKAAKAAGHNITVPTFGRTDASQEQTDVFSFAVLEPKADGFR 595
QY 600 NYIKPEKVSAAEMLVDRAQLLSAPETALYVGMRLVLTNTYDGSQHGVTNKPGLSN 659
Db 596 NYLKGKPLPAEYMLLDKANLLLSAPEMTVLVGLRVLGANYKRLPLGVFTASESLTN 655
QY 660 DFFVNLLDLNTKWRASDESKVEGRDFKTEGVKWSGTRVDLIFGSNSSELRLAELVYGA 719
Db 656 DFFVNLDMGITWEPSPADDTGYGKD-GSGKVKWTGSRVDLVFGSNSSELRLALVEVYGA 714
QY 720 DSEKFKVQFKAWKAWMDLDRDLK 745
Db 715 DAQPKFVQDEVAAMDVMNLDREDFVR 740

```

```

RESULT 5
US-08-418-782-21
; Sequence 21, Application US/08418782
; Patent No. 5658733
; GENERAL INFORMATION:
; APPLICANT: Cockerill, Franklin R.
; APPLICANT: Kline, Bruce C.
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Isoniazid Resistant Strains
; TITLE OF INVENTION: of M. Tuberculosis
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg & Woessner

```

```

; STREET: 3500 IDS Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,782
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D.
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 150.141US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 740 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-418-782-21

```

```

Query Match      63.6%; Score 2544; DB 1; Length 740;
Best Local Similarity 64.3%; Pred. No. 9.5e-222;
Matches 480; Conservative 88; Mismatches 168; Indels 10; Gaps 8;

QY 2 ENHKHSSSYNTNTGKCPPTGSLKQSGAGGKTRNDWPNMLNLGTLQHSLSLSDPND 61
Db 3 EGHPIETTTGAASNG-CPVVGHHKYPVEGG--NQDMPNRLNLKVLHONPAVADPMG 59
QY 62 PDFDYAEFFKDLAAVKDLAALTDSDQWMPADYGHYGFPTFMAHWSAGTYRIGDGR 121
Db 60 AAFDYAAEVATIDVDAITRDIEEVYTTSQPWPADYGHYGLFTRMAHAAAGTYRIHGR 119
QY 122 GGGGSGSORFAPLNSWPNANLDRKLLWPVKYKGRKISWADMLTGNVALETMGFK 181
Db 120 GGAGGGMORFAPLNSWPNASLDKARLLMPVKKYKGLSWADLIYFAGNCALESNGFK 179
QY 182 TFGFAGGRADWPEEDVYGAETWLGDKRYEGDRELENPLGAVQMLIYVNPGEPMGK 241
Db 180 TFGFGRVQWPEDE-VYWGKEATWLGDERYSKRDLENPLAAVQMLIYVNPGEPMGN 238
QY 242 PDPTAAARDIRETFRMAMNDEETVALIAGHTFGKTHGAADAKEYYVGREPAAAGIEMS 301
Db 239 PDPMAAAVDIRETFRMAMNDVETAAALIVGGHTFGKTHGAGPAD-LVGPEPEAAPLEQMG 297
QY 302 LGWKNTYGTGHGADTITSGLEGATWKTPTOWSNFFENLFGYEWELTKSPAGAYQWPKD 361
Db 298 LGWKSSYGTGKDAITSGIEEVVYNTPTKWDNSFLIYGYEWELTKSPAGAWQYTKD 357
QY 362 GAGAGTTPDAH-DPSKSHAPFMTLTDLALRMDPDYKISRYYENPDEFAFAKAWYKL 420
Db 358 GAGAGTTPDPFGGGRS--PTMLATDLSLRVDPYIERITRWLEHPELDEFKAWYKL 415
QY 421 THRDMPKRYLGPVEPQEDLIWQDPIPVSHPLVDENDIEGLKAKILESGLTVELYST 480
Db 416 IHRDMGPVARYLGPLVPKQTLWQDPVAVSHDLVGEAETASLSKQTRASGLTVSQVYST 475
QY 481 AWASASTFRNSDKRGANGARIRLAPQKDWENNPQ-QLARVLKTLLEGIQEDFNQAOSDN 539
Db 476 AWAAASSFRGSKRGANGGRIQLPQVQWGEVNDPDGDLRKVIRTLKEIOESFNSAAGN 535
QY 540 KAVSLADLIYLACAGVEKAKADAGHEVQVFPNPGRADATAEQTDVFAFPALEPAADGFR 599
Db 536 IKVSFADLVVLGGCAAEKAAKAAGHNITVPTFGRTDASQEQTDVFSFAVLEPKADGFR 595

```

```
QY 600 NYKPEHKVSAEMLVDRAQLLSLSAPENTALVGMRLVLTNYDGSQHGVTNKPQGLSN 659
Db 596 NYLGKGNPLPAEYMLLDKANLLTLSAPENTVVLVGLRVLGANYRLPLGVFTSEASESLTN 655
QY 660 DFFVNLDLNTKWRASDESDFVFGROFPGTKTEVXWSGTRVDLIFGNSSELRALAEVYGA 719
Db 656 DFFVNLDMGITWEPSPADDTYQKDGK-GSGKVTWTSRVDLIFGNSSELRALVEVYGAD 714
QY 720 DSEKFKVDFKAWAKVMDLDRFDLK 745
Db 715 DAQPKFVQDFVAAMDVKVNLDRFDVR 740

RESULT 6
US-08-228-662-7
; Sequence 7, Application US/08228662
; Patent No. 5688639
; GENERAL INFORMATION:
; APPLICANT: COCKERILL, FRANKLIN R.
; APPLICANT: KLINE, BRUCE C.
; TITLE OF INVENTION: DETECTION OF ISONIAZID RESISTANT STRAINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCHWEGMAN, LONDBERG & WOESSNER, P.A.
; STREET: 3500 IDS CENTER
; CITY: MINNEAPOLIS
; STATE: MINNESOTA
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/228,662
; FILING DATE: 18-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WOESSNER, WARREN D.
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 150.123US01
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 740 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-228-662-7

Query Match 63.6%; Score 2544; DB 1; Length 740;
Best Local Similarity 64.3%; Pred. NO. 9.5e-222;
Matches 480; Conservative 88; Mismatches 168; Indels 10; Gaps 8;

QY 2 ENHKGSGSTYNTYGGCKPTGGLSKQSAGGKTNRDWWPNMLNLGLILRHSSLSLDPND 61
Db 3 EGHGPTTETTGASNG-CPVVGHKYKPVVEGG--NODWPNRLNLKYLHONPAVADPMG 59
QY 62 PDFDYAEFFKLDLAANKDLAALMTSDQWNPADYGHYGPFFIRMAWHSAGTVRIGDGR 121
Db 60 AAFDYAEVAIDVDALTRDIEEYVWTSQPPWADYGHYGLPFINMAWHAAGTVRIHGR 119
QY 122 GGGGSGGQRFAPLNSWPDNANLDRKLLWPIKQYGRKISWADLMILTNVALETMGFK 181
Db 120 GGAGGGMQRAPLNSWPDNANLDRKLLWPIKQYGRKISWADLMILTNVALETMGFK 179
QY 182 TFGAGGADYWEPEEDYVNGAETEWLGDKKRYEGDRELENPLGAVQVOMGLIYVNPBGNK 241
```

```
Db 180 TFGFGVQDQWEPDE-VYWGKEATWLGDERYSGRKDLLENPLAAVOMGLIYVNPBGN 238
QY 242 PDPLAAARDIRETFGRMAAMDEETVALLAGHTTGTGTHGAADAERYVGVREPAAGAEIEMS 301
Db 239 FDPMAAADYDIRETFGRMAAMNDVETAAVLGVHTFGTHGAGPAD-LVGPPEPAAPLEQMG 297
QY 302 LGWNTYGTGHGADITISGLEGAWTKTPTQMSNNFFENLFGYEWELTKSPAGAYQWKPRD 361
Db 298 LGWSSYGTGTGKDAITSGIEVWVTNTPTKWDNSFLEILYGYEWELTKSPAGAWQYTKD 357
QY 362 GAGAGTIPDAH-DPSKSHAPFMLTTDLALRMDPDYEKISRYYENPDPEADAFKAWYKL 420
Db 358 GAGAGTIPDPFGGPGRS--PTMLATDLSLRVDPIYERITRRLWLEHPEELADEFAKAWYKL 415
QY 421 THRDGPKRVRYLGPVPOEDLIWQDPIPDVSHPLVDENDIEGLKAKILESGLTVSELVST 480
Db 416 IHRDGPVARYLGPLVPKOTLLWQDPVPAVSHDLVGEAEIASLKSQIRASGLTVSQLVST 475
QY 481 AWASASTFRNSDKRGANGARIRLAPQKDWENNPQ-QLARVLTKEGIEQEDFNQAQSDN 539
Db 476 AWAAASSFRGDKRGANGRRIRLQPVQWGVNDPDGLRKVIRTLREETQESFNSAAPGN 535
QY 540 KAVSLADLIVLAGCAGVEKAAKADGAGHEVQVFPNPGRADATAGTQDYEAFALEPADGFR 599
Db 536 IKVSFADLVWLGCGAAIEKAAKAAGHNTVPTPGRTDASQSDTVESFAVLEPKADGFR 595
QY 600 NYKPEHKVSAEMLVDRAQLLSLSAPENTALVGMRLVLTNYDGSQHGVTNKPQGLSN 659
Db 596 NYLGKGNPLPAEYMLLDKANLLTLSAPENTVVLVGLRVLGANYRKRLPLGVFTSEASESLTN 655
QY 660 DFFVNLDLNTKWRASDESDFVFGROFPGTKTEVXWSGTRVDLIFGNSSELRALAEVYGA 719
Db 656 DFFVNLDMGITWEPSPADDTYQKDGK-GSGKVTWTSRVDLIFGNSSELRALVEVYGAD 714
QY 720 DSEKFKVDFKAWAKVMDLDRFDLK 745
Db 715 DAQPKFVQDFVAAMDVKVNLDRFDVR 740

RESULT 7
US-08-852-219-7
; Sequence 7, Application US/08852219
; Patent No. 5922575
; GENERAL INFORMATION:
; APPLICANT: COCKERILL, Franklin R.
; APPLICANT: Kline, Bruce C.
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Isoniazid Resistant Strains
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mueeting, Raasch & Gebhardt, P.A.
; STREET: 119 No. 5922575th Fourth Street, Ste. 203
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,219
; FILING DATE: 07-May-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sandberg, Victoria A.
; REGISTRATION NUMBER: 41,287
; REFERENCE/DOCKET NUMBER: 230.00010130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1226
```

TELEFAX: 612-305-1228
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 740 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-852-219-7

Query Match 63.68; Score 2544; DB 2; Length 740;
 Best Local Similarity 64.38; Pred. No. 9.5e-222;
 Matches 480; Conservative 88; Mismatches 168; Indels 10; Gaps 8;

QY 2 ENKHGSSSYNTNGCKPFTGSLKQSGAGGTKNRDMWPNMLNLGILRQHSLSLSPND 61
 DB 3 EGHPPTTETTTGAASNG-CPVGHMKYPVEGGG--NQDWWPNRLNLKVLHONPAVADPMG 59

QY 62 PDFDYAEFFKKDLAAVKDLAALMTDSQDWNPADYGHYGFPIRMAHWSAGTYRIGDGR 121
 DB 60 AAFDYAAEVATIDVLTDRDIEVMTTSPQWNPADYGHYGFPIRMAHWSAGTYRIGDGR 119

QY 122 GGGGSSQRPAPLNSWPDNANLKPILKQYGRKISWADLMTLTGNVALETMGFK 181
 DB 120 GGAGGGQRPAPLNSWPDNANLKPILKQYGRKISWADLMTLTGNVALETMGFK 179

QY 182 TFGAGRADVPEDVYGAETWLGDKRYSGDRELENPLGAVOMGLIYVNPGEPMG 241
 DB 180 TFGGGRVQWEPDE-VYMGKEATWLGDRYSKGRDLNPLAIVGAGPAD-LVGPPEAPLEQMG 238

QY 242 PDPTAAARDIRETFGRMANDEETVALIAGGHTFGKTHGAADAKEYVGRPAAGTEMS 301
 DB 239 PDPTAAARDIRETFGRMANDEETVALIAGGHTFGKTHGAADAKEYVGRPAAGTEMS 297

QY 302 LGWKSSYGTGKDAITSGLEGAWTKPTQWSNNFENLFGYEWELTKSPAGAWQYAKD 361
 DB 298 LGWKSSYGTGKDAITSGLEGAWTKPTQWSNNFENLFGYEWELTKSPAGAWQYAKD 357

QY 362 GAGAGTIPDAH-DPSKSHAPFMTLTDLALRMDPDYKISRRYENPDEFAFAKAWYKL 420
 DB 358 GAGAGTIPDPFGGPGRS--PTMLATDLSLRVDPIYERITRWLEHPEELADEFAKAWYKL 415

QY 421 THRDMPKRYLGPVEPQEDLIWQDPIPDVSHPLVDENDIEGLKAKILESGLTVELYST 480
 DB 416 IHRDMGPVARYLGPLVKQTLWQDPPVAVSHDLVGEAEIASLKSQIRASGLTVSGLYST 475

QY 481 AWASASTFRNSDRGGANGARIRLAPQKDWENNPQ-QLARVLKTLGIEQEDFNQAOSDN 539
 DB 476 AWASASTFRNSDRGGANGARIRLAPQKDWENNPQ-QLARVLKTLGIEQEDFNQAOSDN 535

QY 540 KAVSLADLIVLACGACGAKAAGHEVQVPPNPGRADATATQDTVEAFALPEADGFR 599
 DB 536 IKYSFADLVVLGCAATEKAAKAAAGHNITVPPTGRDASQEQTDVESFAVLEPKADGFR 595

QY 600 NYLKPEHVSAREMLVDRAOLLSAPENTALVGGMRVLGTNYDGGSHGVFTKPKQLSN 659
 DB 596 NYLKGNPLPAETMLDKNLLTSLAPENTVVLGGRLVGLGANTKRLPLGVFTTASPSLTN 655

QY 660 DFFVNLDLTKWRASDESQVPEGRDFKTEGKWSGTRVDLIFGNSSELRLAELAYGCA 719
 DB 656 DFFVNLDMGITEWPSPADDTGQKGD-GSGKVKWTGSRVDLVFGNSSELRLALVEYIGAD 714

QY 720 DSEKPKVQKPAKAWKVMOLDREPLK 745
 DB 715 DAQPKFVQDPVAAWDKVMNLDREPR 740

RESULT 8
 US-08-852-219-21
 Sequence 21, Application US/08852219
 Patent No. 5922575
 GENERAL INFORMATION:
 APPLICANT: Cockerill, Franklin R.

APPLICANT: Kline, Bruce C.
 APPLICANT: Uhl, James R.
 TITLE OF INVENTION: Detection of Isoniazid Resistant Strains
 TITLE OF INVENTION: of M. Tuberculosis
 NUMBER OF SEQUENCES: 22
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Muetling, Raasch & Gebhardt, P.A.
 STREET: 119 No. 5922575th Fourth Street, Ste. 203
 CITY: Minneapolis
 STATE: MN
 COUNTRY: USA
 ZIP: 55401

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/852,219
 FILING DATE: 07-May-1997
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Sandberg, Victoria A.
 REGISTRATION NUMBER: 41,287
 REFERENCE/DOCKET NUMBER: 230.00010130
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 612-305-1226
 TELEFAX: 612-305-1228
 INFORMATION FOR SEQ ID NO: 21:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 740 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-852-219-21

Query Match 63.68; Score 2544; DB 2; Length 740;
 Best Local Similarity 64.38; Pred. No. 9.5e-222;
 Matches 480; Conservative 88; Mismatches 168; Indels 10; Gaps 8;

QY 2 ENKHGSSSYNTNGCKPFTGSLKQSGAGGTKNRDMWPNMLNLGILRQHSLSLSPND 61
 DB 3 EGHPPTTETTTGAASNG-CPVGHMKYPVEGGG--NQDWWPNRLNLKVLHONPAVADPMG 59

QY 62 PDFDYAEFFKKDLAAVKDLAALMTDSQDWNPADYGHYGFPIRMAHWSAGTYRIGDGR 121
 DB 60 AAFDYAAEVATIDVLTDRDIEVMTTSPQWNPADYGHYGFPIRMAHWSAGTYRIGDGR 119

QY 122 GGGGSSQRPAPLNSWPDNANLKPILKQYGRKISWADLMTLTGNVALETMGFK 181
 DB 120 GGAGGGQRPAPLNSWPDNANLKPILKQYGRKISWADLMTLTGNVALETMGFK 179

QY 182 TFGAGRADVPEDVYGAETWLGDKRYSGDRELENPLGAVOMGLIYVNPGEPMG 241
 DB 180 TFGGGRVQWEPDE-VYMGKEATWLGDRYSKGRDLNPLAIVGAGPAD-LVGPPEAPLEQMG 238

QY 242 PDPTAAARDIRETFGRMANDEETVALIAGGHTFGKTHGAADAKEYVGRPAAGTEMS 301
 DB 239 PDPTAAARDIRETFGRMANDEETVALIAGGHTFGKTHGAADAKEYVGRPAAGTEMS 297

QY 302 LGWKSSYGTGKDAITSGLEGAWTKPTQWSNNFENLFGYEWELTKSPAGAWQYAKD 361
 DB 298 LGWKSSYGTGKDAITSGLEGAWTKPTQWSNNFENLFGYEWELTKSPAGAWQYAKD 357

QY 362 GAGAGTIPDAH-DPSKSHAPFMTLTDLALRMDPDYKISRRYENPDEFAFAKAWYKL 420
 DB 358 GAGAGTIPDPFGGPGRS--PTMLATDLSLRVDPIYERITRWLEHPEELADEFAKAWYKL 415

QY 421 THRDMPKRYLGPVEPQEDLIWQDPIPDVSHPLVDENDIEGLKAKILESGLTVELYST 480
 DB 416 IHRDMGPVARYLGPLVKQTLWQDPPVAVSHDLVGEAEIASLKSQIRASGLTVSGLYST 475

QY 481 AWASASTFRNSDRGGANGARIRLAPQKDWENNPQ-QLARVLKTLGIEQEDFNQAOSDN 539

Db 476 AWAASFRGSKRGANGRRIRLOPQVGEVNDPDGDLRKVIRLIEIODESFAAPGN 535
Qy 540 KAVSLADLIVLACAGVEKAADAGHEVQVFPNPRADATAEQDVEAFEALEPAADGFR 599
Db 536 IKVSFADLVLGCGAAIEKAAGAAGNITVPTFGRTDASQEDVDFEFAVLEPAADGFR 595
Qy 600 NYIKPEHKVSAEMLVDRALQILSLASAPEMTALVCGMRVLGTYNDCSHGQVFTNPKQOLSN 659
Db 596 NYLKGKGNPLPAEYMLLDKANLLTSLAPEMTVLVGLRVLGANGKRLPLGVFTEASESILTN 655
Qy 660 DFFVNLDLNTKWRASDESCKVPRGDKTGEVKNWSTGRVDLIFGNSSELRALAEVYGA 719
Db 656 DFFVNLDMGITWEPSPADDTGYGKD-GSGKVAWTSRVDLVFGNSSELRALVEVYGA 714
Qy 720 DSEKFKYKDFYKAWKYMNDLDRDLK 745
Db 715 DAQPKFQVDFVAAMDVKVNNLDRPDVR 740

RESULT 9
US-08-313-185-49
; Sequence 49, Application US/08313185
; Patent No. 5851763
; GENERAL INFORMATION:
; APPLICANT: Heym, Beate
; APPLICANT: Cole, Stewart
; APPLICANT: Young, Douglas
; APPLICANT: Zhang, Ying
; APPLICANT: Honore, Nadine
; APPLICANT: Telenti, Amalio
; APPLICANT: Bodmer, Thomas
; TITLE OF INVENTION: Rapid Detection of Antibiotic Resistance
; TITLE OF INVENTION: in Mycobacterium Tuberculosis
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313,185
; FILING DATE: 12-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 02356.0068-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 726 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-313-185-49

Query Match 60.5%; Score 2421.5; DB 2; Length 736;
Best Local Similarity 63.6%; Pred. No. 1.2e-210;
Matches 467; Conservative 78; Mismatches 172; Indels 17; Gaps 7;
Qy 12 YNTNTGGKCPFTGGSLKOSAGGTKNRDWWPNMLNLGILRHSLSLSDPNDFDYAEBFK 71

Db 8 HNTTATGKCPHQGHDSAGAGTTTRDWWPNQLRVLLNQHSNRSNPLGDFDYRKEFS 67
Qy 72 KLDLAAVKDLAALMTDSQDWWPADYGHYGFPFFIRMAHWSAGTYRIGRGGGSGSORF 131
Db 68 KLDYIGLKKDLKALLTSQPPWADWSYAGLFRMAHWSAGTYRIGRGGGSGSORF 127
Qy 132 APLNSWPDNANLDRKALLWPKYKGRKISWADLMILTGNALETMGFTKTFGAGGRAD 191
Db 128 APLNSWPDNANLDRKALLWPKYKGRKISWADLMILTGNALETMGFTKTFGAGGRAD 187
Qy 192 VWEPEEDVYCAETFWLGDCKRYEGDRELENLGAQVOMGLIYVNPGEKPKDPTAAARDI 251
Db 188 VWEPEEDVYCAETFWLGDCKRYEGDRELENLGAQVOMGLIYVNPGEKPKDPTAAARDI 245
Qy 252 RETEGRMANDEETVALIAGHTGKTHGAADAKEYVGRPEAAAGIEMSLGKWNKTYGTG 311
Db 246 RATEGNMCMNDEETVALIAGHTGKTHGAAGTSTN-VGPDPEAPIEQGLGWASTYSG 304
Qy 312 HGADTITSGLEGANTKPTTOWSNFFENLFGYEWELTKSPAGAYQWPKDGAGATIPDA 371
Db 305 VGADAITSGLEVVWTTPTOWSNFFENLFGYEWELTKSPAGAYQWPKDGAGATIPDA 362
Qy 372 HPSKSHAPFMTTDLALRMDPDYKISRYYENPDEPADAFAKAWYKLTIRDMGPKRY 431
Db 363 FDPKSKRPTMLVTDLTURFDFEIKSRRLNDPQAFNEAFARAWFKLTIRDMGPKRY 422
Qy 432 LGPEVPOEDLIWQDPIPD-VSHPLVDNDIEGLKAKILESGLTVSELYSTAWASSTFRN 490
Db 423 IGPEVPKEDLIWQDPLQPIYNP--TEQDIIDLKFAIDSLVSELVSVAVASSTFRG 480
Qy 491 SDRKGGANGARIRLAPQKQDEVNPPQQLARVLKLEGIQEDFNQASDNKAVSLADLIVL 550
Db 481 GDRKGGANGARIRLAPQKQDEVNPPQQLARVLKLEGIQEDFNQASDNKAVSLADLIVL 531
Qy 551 ACCAGVEKAADAGHEVQVFPNPRADATAEQDVEAFEALEPAADGFRNTIKPKHKVSA 610
Db 532 AGVVGVEKAASAAGLSIHVPFAPGRVDARQDQTDIEMFELLEPIADGFRNTRARLDVSTT 591
Qy 611 EEMLVDRQAQLLSLAPENTLVGMRVLGTYNDCSHGVTNKPQLSNDFFVNLDLNT 670
Db 592 ESLLDRAQQLTLTAPENTLVGMRVLGTYNDCSHGVTNKPQLSNDFFVNLDLNT 651
Qy 671 KWRASDESCKVPRGDKTGEVKNWSTGRVDLIFGNSSELRALAEVYGADEEKFKVDFV 730
Db 652 EWKATDESCKVPRGDKTGEVKNWSTGRVDLIFGNSSELRALAEVYGADEEKFKVDFV 711
Qy 731 KAWKVMDLDRFDL 744
Db 712 AAWKVMNLDRFDL 725

RESULT 10
US-08-459-499-13
; Sequence 13, Application US/08459499
; Patent No. 5871912
; GENERAL INFORMATION:
; APPLICANT: Heym, Beate
; APPLICANT: Cole, Stewart T.
; APPLICANT: Young, Douglas B.
; APPLICANT: Zhang, Ying
; TITLE OF INVENTION: Nucleic Acid Probes, Sequences, and Methods
; TITLE OF INVENTION: for Detecting Mycobacterium Tuberculosis Resistant to Isoni
; TITLE OF INVENTION: Amended)
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315

Db 590 DVSTTESLLIDKAQQLTLTAPETVLVGGMRVLGTFNFGSQNGVFTDRPGVLTDFANL 649
Qy 666 LDLNTKWRASDESKVFEGRDFKTEVGMSTGRVLDLFGSNSELRALAEVYGCADSEK 725
Db 650 LDMRYENKPTDANELPEGRDLTGCVKYTTATRADLVFGSNVLRALAEVYACSDAHEK 709
Qy 726 VKDFVKAQVMDLDRDLK 745
Db 710 VKDFVAAWVKNLDRDLQ 729

RESULT 13
US-08-459-499-14
; Sequence 14, Application US/08459499
; Patent No. 5871912
; GENERAL INFORMATION:
; APPLICANT: Heym, Beate
; APPLICANT: Cole, Stewart T.
; APPLICANT: Young, Douglas B.
; APPLICANT: Zhang, Ying
; TITLE OF INVENTION: Nucleic Acid Probes, Sequences, and Methods
; TITLE OF INVENTION: for Detecting Mycobacterium Tuberculosis Resistant to Isoniazid
; TITLE OF INVENTION: Amended
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.3
; CURRENT APPLICATION DATA: US/08/459,499
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/875,940
; FILING DATE: 30-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/929,206
; FILING DATE: 27-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/029,655
; FILING DATE: 11-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0110-03000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 729 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-459-499-14

Query Match 59.9%; Score 2397; DB 2; Length 729;
Best Local Similarity 62.8%; Pred. No. 2e-208;
Matches 465; Conservative 82; Mismatches 171; Indels 22; Gaps 9;

Qy 11 TYNTWCGKCPFTGGSLKQAGGCTKRNDRWPNMNLGILRQHSLSLSDPNPDYAEF 70
Db 7 THNTLSTGKCPFHGGHDSAGAGTASGWPNLRLVLLNQHNSRNLPLGEDFDYRKEF 66

Qy 71 KKLD-LAAVKKDLAALMTDSQDWMPADYGHYGPFFIRMAHWSAGTYRIGDGRGGGSSQ 129
Db 67 SKLDYYSALKGLKALLTDSQDWMPADYGHYGPFFIRMAHWSAGTYRIGDGRGGGSSQ 126
Qy 130 RFAPLNSPDNANLKDARLLWPIKQYGRKISWADLMLTGNVALETWGTGTFPAGGR 189
Db 127 RFAPLNSPDNANLKDARLLWPIKQYGRKISWADLMLTGNVALETWGTGTFPAGGR 186
Qy 190 ADVWEPEEDVYGAETEMVGLDKRYBGDRELENPICAVQMGVLYVNPENGPNGKDPDIAAAR 249
Db 187 EDVWEPEEDVYGAETEMVGLDKRYBGDRELENPICAVQMGVLYVNPENGPNGKDPDIAAAR 244
Qy 250 DIRETFGRMANDEETVALIAGGHTFGTHGAADAKEYVGRPEAAAGIEEMSLGKNTYG 309
Db 245 AIRATFGNMGNDDEETVALIAGGHTFGTHGAADAKEYVGRPEAAAGIEEMSLGKNTYG 303
Qy 310 TGHGADTITSGLEGATWTKPTOWSNFFENLPGYEWELTKSPAGAYQKPKDGAGATIP 369
Db 304 SGVGADAITSGLEGATWTKPTOWSNFFENLPGYEWELTKSPAGAYQKPKDGAGATIP 361
Qy 370 DAHPDSKSHA--PFMLTTLALRMDPDYKISRRYVENPDEFADAFKAWKLTTHRMGP 427
Db 362 DPDPSSKRRXXKPTMLVTDLTLRDPDEFEKISRRLNDPQAFNEAFARAWKLTTHRMGP 421
Qy 428 KRYLGPVPOEDLIWQPIPDVSHPLVD--ENDIEGLKAKILESGLTVSELVSTAWASA 485
Db 422 KARYIGPEVPKEDLIWQPIPDVSHPLVD--ENDIEGLKAKILESGLTVSELVSTAWASA 478
Qy 486 STERNDRKGGANGARIRLAPQKQWENNPQOLARVLKTLGSIQSDNQAQSDNKAWSLA 545
Db 479 STFRGGDRGGANGARIRLAPQKQWENNPQOLARVLKTLGSIQSDNQAQSDNKAWSLA 529
Qy 546 DLIVLAGCAGVEKAAKADAGHEVQVFPNPGRADATAEQTDDVEAFEALEPAADGFRNYIKPE 605
Db 530 DIIVLAGVVGTEQAAAAARVSIHVPFPGVGRVARDHDDTDIEMFSLLEPIADGFRNYIKPE 589
Qy 606 HKVSAEMLVDRAQLLSAPEMTALVGGMRVLGTFNFGSQNGVFTDRPGVLTDFANL 649
Db 590 DVSTTESLLIDKAQQLTLTAPETVLVGGMRVLGTFNFGSQNGVFTDRPGVLTDFANL 649
Qy 666 LDLNTKWRASDESKVFEGRDFKTEVGMSTGRVLDLFGSNSELRALAEVYGCADSEK 725
Db 650 LDMRYENKPTDANELPEGRDLTGCVKYTTATRADLVFGSNVLRALAEVYACSDAHEK 709
Qy 726 VKDFVKAQVMDLDRDLK 745
Db 710 VKDFVAAWVKNLDRDLQ 729

RESULT 14
US-09-082-614A-50
; Sequence 50, Application US/09082614A
; Patent No. 6124098
; GENERAL INFORMATION:
; APPLICANT: Heym, Beate
; APPLICANT: Cole, Stewart
; APPLICANT: Young, Douglas
; APPLICANT: Zhang, Ying
; APPLICANT: Honore, Nadine
; APPLICANT: Telenti, Amelio
; APPLICANT: Bodmer, Thomas
; TITLE OF INVENTION: Rapid Detection of Antibiotic Resistance
; TITLE OF INVENTION: in Mycobacterium Tuberculosis
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:

QY	617	RAQLLSAPENTALVCGMRVLGTYDGSQHGVTNKPQQLSNDFFVYNLLDLNTKWRASD	676
Db	598	KAQQLTLTAPELTALIGLRLVLTGTYDGSQHGVTNKPQQLSNDFFVYNLLDLNTKWRASD	657
QY	677	ESDKVFEGRDFTKGEVKWSGTRVDLIFGSNSLRALAEVYGCADSEKFKVDFVKAQKV	736
Db	658	STSEVFEGRDFTKGEVKWSGTRVDLIFGSNSLRALAEVYGCADSEKFKVDFVKAQKV	717
QY	737	MDLDREFDL 744	
Db	718	MNLDREFDL 725	

Search completed: October 7, 2003, 19:47:37
Job time : 24.8825 secs

1	Qy	MENHHSGSSTYNTNTGGKCPFTTGGSLKOSAGGTTNRDWWPNMLGLILRQHSSLSDPN	60
1	Db	MENHHSGSSTYNTNTGGKCPFTTGGSLKOSAGGTTNRDWWPNMLGLILRQHSSLSDPN	60
61	Qy	DDPDYAEEFKKLIDLAARVKDLAALMTDSQDWMPADYGHYGPFFIRMAVHSACTYRI GDG	120
61	Db	DDPDYAEEFKKIYLAARVKDLAALMTDSQDWMPADYGHYGPFFIRMAVHSACTYRI GDG	120

QY 121 RGGGGGSGRFPAPLNSWPDNANLDKARLLNLPKQYGRKISWADLMILTGNVALETMGF 180
Db |||||
QY 121 RGGGGGSGRFPAPLNSWPDNANLDKARLLNLPKQYGRKISWADLMILTGNVALETMGF 180
Db |||||
QY 181 KTFGAGGADYWEPEEDYWGAEETWMLGDKRYEGDRELENPLGAVQMGGLIYVNEGPN 240
Db |||||
QY 181 KTFGAGGADYWEPEEDYWGAEETWMLGDKRYEGDRELENPLGAVQMGGLIYVNEGPN 240
Db |||||
QY 241 KPDPTAAARDIRETGRMANDEETVALIAGHTFGKTHGAADAERYGREGPAAGIEM 300
Db |||||
QY 241 KPDPTAAARDIRETGRMANDEETVALIAGHTFGKTHGAADAERYGREGPAAGIEM 300
Db |||||
QY 301 SLGWKNTGTGAGADITISGLEGAWTKPTQWNSNFFENLFGYEWELTKSPAGAYQWPK 360
Db |||||
QY 361 DGAGAGTIPDAHDPSKSHAPFMTTDLALRMDPDYKISRRYYENPDEFADAFARAWYKL 420
Db |||||
QY 421 THRDMPKRYLGPVPOEDLIWQDPIPDVSHPLVDENDIEGLKAKILESGLTVSELYST 480
Db |||||
QY 481 AWASASTFRNSDKRGANGARIRLAPQKDWENPNPQOLARVLKTLGEOEDFNQAQSDNK 540
Db |||||
QY 541 AVSLADLIVLAGCAGVKAADAGHEVQVPPNPGRADATAEQTQDVEAFEALEPAADGPRN 600
Db |||||
QY 601 YIKPEHKVSAEMLVDRAQLLSAPEMTALVGGMRVLGTNYDSQHGVTNKPQLSND 660
Db |||||
QY 661 FVNLDDLNTKWRASDESCKVFEGRDFTKGEVKSQGTVDLIFGNSNSELRALEAYYGAD 720
Db |||||
QY 721 SEERFVKDFVKAQVMDLDRFDLK 745
Db |||||

RESULT 2

US-09-884-889-6
; Sequence 6, Application US/09884889
; Patent No. US20020102680A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: ROBERTSON, Dan
; APPLICANT: SANYAL, Indrajit
; APPLICANT: ADHIKARI, Robert
; TITLE OF INVENTION: CATALASES
; FILE REFERENCE: DIVER1100-4
; CURRENT APPLICATION NUMBER: US/09/884,889
; CURRENT FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 09/412,347
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: US 08/951,844
; PRIOR FILING DATE: 1997-10-16
; PRIOR APPLICATION NUMBER: US 08/674,887
; PRIOR FILING DATE: 1996-07-03
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 753
; TYPE: PRF
; ORGANISM: Alcaligenes (Deleya) aquamarinus
US-09-884-889-6

Query Match 57.2%; Score 2290.5; DB 10; Length 753;
Best Local Similarity 59.1%; Pred. No. 2.1e-200;
Matches 433; Conservative 92; Mismatches 183; Indels 25; Gaps 8;
QY 18 GKCPFTGSSLKASAGGATKNRDWPMLNGLIRQHUSLSDDNDPDDFYAEFFKLDLAA 77
Db |||||
QY 38 GKCPVWHGG---NTSGTSSKNDWPEGLNLDLHQODRKSDDPMDPNDYREVRKLDLDA 94
Db |||||
QY 78 VKKDAAALMTDSQDWMPADYGHYGPFFIRMAHSAAGTYRIGDGRGGSGSRRFAPLNSW 137
Db |||||
QY 95 LAKDVEALMTDSQEWMPADWGHYGLMIRMAHSAAGTYRIADRGGGGTGSGRRFAPLNSW 154
Db |||||
QY 138 PONANLDKARLLNLPKQYGRKISWADLMILTGNVALETMGKTFGAGGADYWEPE 197
Db |||||
QY 155 PONVSLDKARLLNLPKQYGRKISWADLMILTGNVALETMGKTFGAGGADYWEPE 214
Db |||||
QY 198 DYTWGAETEWL--GDKRYEGD---RELENPLGAVQMGGLIYVNEGPNKPDIAAARDI 251
Db |||||
QY 215 DIYWGDEKEWLAAPSERY--GDVKNKPTMBENPLAAYQMGGLIYVNEGPNHDPDLRTAQOV 273
Db |||||
QY 252 RETFRMANDEETVALIAGHTFGKTHGAADAERYGREGPAAGIEMSLGKWKNTYGTG 311
Db |||||
QY 274 LETFRMANDEETVALIAGHTFGKTHGAADAERYGREGPAAGIEMSLGKWKNTYGTG 332
Db |||||
QY 312 HGADTITSGLEGAWTKPTQWNSNFFENLFGYEWELTKSPAGAYQWPKDAGAGACTIPDA 371
Db |||||
QY 333 KASNAVTSIGEGAWTNTPTKEDMGYFDLLFGYNWELKSPAGAHWEPIIDIKKENKPYDA 392
Db |||||
QY 372 HDPSKSHAPFMTTDLALRMDPDYKISRRYYENPDEFADAFARAWYKLTHRDMPKRY 431
Db |||||
QY 393 SDPSIRHNPIMTDADMAIKVNPYRAICEKFMADPEYFKTFPAKAWFKLTHRDLPKSY 452
Db |||||
QY 432 LGPEVPOEDLIWQDPIPDVSHPLVDENDIEGLKAKILESGLTVSELVSTAWASASTFRNS 491
Db |||||
QY 453 IGPEVPAEDLIWQDPIPDVSHPLVDENDIEGLKAKILESGLTVSELVSTAWASASTFRNS 508
Db |||||
QY 492 DKRGANGARIRLAPQKDWENPNPQOLARVLKTLGEOEDFNQAQSDNKAVSLADLIVLA 551
Db |||||
QY 509 DMRGANGARIRLAPQKDWENPNPQOLARVLKTLGEOEDFNQAQSDNKAVSLADLIVLA 560
Db |||||
QY 552 GCAGYKAAKADAGHEVQVPPNPGRADATAEQTQDVEAFEALEPAADGFRNYIKPEHKVSAE 611
Db |||||
QY 561 GSVGLEKAAKAGYDVRVPFLKRGDADAEQTDADSDAPLEPLADGFRNWKKEVYVVKPE 620
Db |||||
QY 612 EMLYDRAQLLSAPEMTALVGGMRVLGTNTYDSQHGVTNKPQLSNDFFVNLDDLNTK 671
Db |||||
QY 621 EMLYDRAQLLSAPEMTALVGGMRVLGTNTYDSQHGVTNKPQLSNDFFVNLDDLNTK 680
Db |||||
QY 672 WRASDESCKVFEGRDFTKGEVKSQGTVDLIFGNSNSELRALEAYYGADSEKFKVDFVK 731
Db |||||
QY 681 WK--PVGSNAYEIRDRKTGAVKWTASRVLDLVFGSNLSLRSAEYVAQDDNGEKFVRDVA 738
Db |||||
QY 732 AWAKVMDLDRFDL 744
Db |||||
QY 739 AWTVMNADRFV 751
Db |||||

RESULT 3

US-09-870-501-1
; Sequence 1, Application US/09870501
; Patent No. US20020100081A1
; GENERAL INFORMATION:
; APPLICANT: Takabe, Tetsuko
; TITLE OF INVENTION: Peroxisomal Ascorbate Peroxidase Gene Induced by High
; TITLE OF INVENTION: Temperature Stress and a Transgeneic Plant Exhibiting
; TITLE OF INVENTION: Thermotolerance
; FILE REFERENCE: 026350-053
; CURRENT APPLICATION NUMBER: US/09/870,501
; CURRENT FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: JP 2000-172850
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1

```
; SEQ ID NO 1
; LENGTH: 291
; TYPE: PRT
; ORGANISM: Hordeum vulgare L. Haruna-niyo
US-09-870-501-1

Query Match
Best Local Similarity 6.8%; Score 273.5; DB 10; Length 291;
Matches 92; Conservative 39; Mismatches 95; Indels 115; Gaps 10;

QY 102 PFFIRMAHSACTYRIGDGRGGSGSORFAPLNSWPNANIDKARLLIWPDKOKYGRKI 161
DQ 32 PIMLRALWHDAGTYDV-NRTGGANGSIRYEEYTHGSGNAGLKIAIDLLLEPIKAH-PKI 89
QY 162 SHADLMILTNVALMETGKTFGAGGRADWPEEDVYWGAEWLGDKRYEGDRELEN 221
DQ 90 TYADLHQLAGVAVVETGPTVEFIPGRD----- 119
QY 222 PLGAVQMGILLYNPESPNGK-PDPIAAARDIRETFGRMAMDEETVALIAGHTGKTHG 280
DQ 120 -----SSVCPR--EGRLEDAKKGAPHLRDIYRMGLTDKDIVAL-SGHSIGKAHP 167
QY 281 ADAEKYVGREPAAAGIEMSLGNWNTYCTGHCADTITSGLEGAWTKPTQWNNFFENL 340
DQ 168 E-----RSGFDGANTROPLKFDNSYF--- 188
QY 341 FGWEWELTKSPAGAYQWPKDGAGAGATIPDAHDPSKSHAPFMTLTDLALRMDPDYEKISR 400
DQ 189 -----LELLK-----GESEGLLKLPDCKALLDDPEFRRYVE 219
QY 401 RYVENPDEFADAFANAKYKLTHTRDMPKVRYLGPVPOEDL 441
DQ 220 LYAKDEDVFFKDYAESHKKLS--ELGFTPRSSGPASTKSDV 258

RESULT 4
US-10-027-559-6
; Sequence 6, Application US/10027559
; Publication No. US20020144307A1
; GENERAL INFORMATION:
; APPLICANT: Simmons, Carl
; TITLE OF INVENTION: Maize Defense-Inducible Genes and Their
; FILE REFERENCE: 035718/239835
; CURRENT APPLICATION NUMBER: US/10/027,559
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 60/243,120
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Zea mays
US-10-027-559-6

Query Match
Best Local Similarity 6.4%; Score 256; DB 14; Length 250;
Matches 88; Conservative 45; Mismatches 99; Indels 126; Gaps 12;

QY 65 DYAEFEKKLDLAAYVKDLAALMTDSQDWPADYGHYGPFFIRMAHSACTYRIGDGRGG 124
DQ 11 EYSEAVDK-----ARRKRLALAEKS-----CAPLMRLAWHSACTFDV-SSRTGG 55
QY 125 GSGSORFAPLNSWPNANIDKARLLIWPDKOKYGRKISWADLMILTNVALMETGKTFG 184
DQ 56 PFGTWKHQSELAHGANAGLDIAVRLLEPIKEEF-PILSTADFYQLAGVAVVETGPGPEIP 114
QY 185 FAGGRADWPEEDVYWGAEWELGDKRYEGDRELENPLGAVQMGILLYNPESPNGKPPD 244
DQ 115 FHPGREDKQPP-----PEG--RLPDA 134
QY 245 IAAARDIRETFGR-MAMNDEETVALIAGHTGKTHGAADAEEKYVGREPAAAGIEMSIG 303

; SEQ ID NO 135
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Hordeum vulgare L. Haruna-niyo
US-09-870-501-1

Query Match
Best Local Similarity 5.7%; Score 229.5; DB 12; Length 287;
Matches 84; Conservative 46; Mismatches 109; Indels 127; Gaps 11;

QY 63 DFYAEFEKKLDLAAYVKDLAALMTDSQDWPADYGHYGPFFIRMAHSACTYRIGDGRG 122
DQ 7 DAELKLEITK-----ARRELSLIANK-----NCAPIMLRALWHDAGTYD-AQSKT 51
QY 123 GGSQSORFAPLNSWPNANIDKARLLIWPDKOKYGRKISWADLMILTNVALMETGKTF 182
DQ 52 GGPNGSIRNEEHTHGANSGLKIALDLCEGVKAH-PKITADLYQLAGVAVVETGPGD 110
QY 183 FGFAGGRADWPEEDVYWGAEWELGDKRYEGDRELENPLGAVQMGILLYNPESPNGK- 241
DQ 111 IVFVPGKD----- 129
QY 242 PDPIAAARDIRETFGRMAMNDEETVALIAGHTGKTHGAADAEEKYVGREPAAAGIEMS 301
DQ 130 PDAKOGFQHLRDVYFVRMGLSDKDIVAL-SGGHTLGRAHPE----- 168
QY 302 LGWKNTYTGHCADTITSGLEGAWTKPTQWNNFFENLFGYEWELTKSPAGAYQWPKD 361
DQ 169 -----RSGFDGPTQEPFLNFDNSYFVR-----ELLK----- 194
QY 362 GAGACTIPDAHDPSKSHAPFMTLTDLALRMDPDYEKISRYYENPDEEFADAFANAKYKLT 421
DQ 195 -----GESEGLLKLPDCKALLDDPEFRRLV-ELYADEDAFFRDYAESHKKLS 240
QY 422 HRDMGP 427
DQ 241 ELGFNP 246

RESULT 6
US-09-734-017A-54
; Sequence 54, Application US/09734017A
; Patent No. US2002014242A1
; GENERAL INFORMATION:
; APPLICANT: Lerchl, Jens
```

```
; APPLICANT: Benz, Andreas
; APPLICANT: Ehrhardt, Thomas
; APPLICANT: Reindl, Andreas
; APPLICANT: Cirpus, Petra
; APPLICANT: Bischoff, Friedrich
; APPLICANT: Frank, Markus
; APPLICANT: Freund, Annette
; APPLICANT: Duwenig, Elke
; APPLICANT: Schmidt, Ralf-Michael
; TITLE OF INVENTION: Moss genes from Physcomitrella patens encoding proteins involved
; TITLE OF INVENTION: the
; TITLE OF INVENTION: synthesis of amino acids, vitamins, cofactors, nucleotides and
; TITLE OF INVENTION: nucleosides
; FILE REFERENCE: BASF-NAE-1331-99-US
; CURRENT APPLICATION NUMBER: US/09/734,017A
; PRIOR FILING DATE: 2000-12-12
; PRIOR FILING DATE: 2000-12-12
; PRIOR FILING DATE: 1999-12-10
; PRIOR FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn Ver. 2.1/WordPerfect
; SEQ ID NO 54
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Physcomitrella patens
; US-09-734-017A-54

Query Match          4.0%; Score 161; DB 10; Length 176;
Best Local Similarity 30.2%; Pred. No. 1.7e-06;
Matches 54; Conservative 22; Mismatches 59; Indels 44; Gaps 7;

Qy 102 PFFIRMAHSAAGYRIGDGRGGGSGSORPAPLNSPNDANLDKARLLLPKQYGRKI 161
Db 34 PTLRLAHSGGYD-OESKTGGPLGTIRFGQELAHGANAGLDIAVNLQPIKEQF-PEL 91
Qy 162 SWADLMILGTNAVALETGFTFGAGGRADWPEEDVYGAETWGLDKRYEGDRELEN 221
Db 92 SYADFYLAVGAVETVGGTTFPHGRKD-----HET-----CPVEG 129
Qy 222 PLGAVOMGLIYNPEGNPKGPDPIAARDIRETFR-MAMNDEETVALIAGGHTFGKTH 279
Db 130 RL-----PDATKGLDLHRCVFTKQMLTDRKDIIVVL-SGAHTLGRCH 169
```

```
RESULT 7
US-09-738-626-6164
; Sequence 6164, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6164
; LENGTH: 867
; TYPE: PRT
```

```
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-6164

Query Match          3.0%; Score 121; DB 10; Length 867;
Best Local Similarity 20.0%; Pred. No. 0.097;
Matches 152; Conservative 94; Mismatches 271; Indels 242; Gaps 42;

Qy 39 DWWPNMLNLGILRQHSSLDPNPDFDYAEFEFKKLDLAAYVKDLAA--LMTDSQ---DWW 93
Db 209 DVWK-----GTLTHHA--ETPAQPTELTVPLALYCRSSSLAKDLDAVRLTETKQGDWY 261
Qy 94 PADYGHYGGP-----FIRMAHSAAGYRIGDGRGGGSGSORPAPLNSPNDAN 142
Db 262 HRNFGVAYPFGKYDQIFVPEFNAGAMENAGAVTIRDEYVFASKATRYE----- 311
Qy 143 LDKARLLLPKQK-YGRKIS--WADLMILGTNAVALETGFTFGAGGRADWPEED 198
Db 312 -RRAETILHELAHMWFGLVTMOMWDDLWL-----NESFATWSAAISOAE--ETEYN 360
Qy 199 VYW----GAETWGLDKRYEGDRELENPLGAVOMGLIYNPEGNPKGPDPIAARDIRET 254
Db 361 TAWVTFANVEKSN---AYQDDQ-----LPSTHPVSDG----- 390
Qy 255 FGRMAMNDEETVALIAGGHTFGKTHGAA----DAEKYVGRPAPAAAGIEE--MSLGWKNTYG 309
Db 391 -----YDIETVDONFDGITYAK--GASVLKQLQAYVGREELAGVRRHFANHAAGNA-- 440
Qy 310 TGHGADYITSGLEGATKTKPTOMSNFFE---NLFCYE----- 344
Db 441 ---SFDLLGALQSSGSRDLSDANQWLKTYGINTLGAKFTTNGKYSFSTVOTGAAPG 497
Qy 345 -----WELTKSPAGAYQWKPKDGAGAGTIPDAHPDSKSHAPMLTTD----- 386
Db 498 AGELRTHRIAVGLYKLVDSGLNRYARVELDCSGASTSVE-EIVGLEQADFVLVNDLTY 556
Qy 387 LALBNPDYKISRRY-YENPDEFAD-----AFKAWYKLTNRDMGPKRYILGPEVPQE 439
Db 557 ALLDLODD-----SRNFVIDNIDKFSDPMPRTLWMSAAWENTRAGOM--KAR----- 601
Qy 440 DLIWQDPIDVSHPLVDENDIEGLKAKILKESGLTVSELVSTAWASASTFRNSDKRGCA-- 497
Db 602 -----DFIALVARGAAAEIEIAVLERILAQATSAKSYADPANAET--GNDLLADFL 653
Qy 498 NGARIRLAPOKDWEVNNPQOLARVLKTLGIEQDFNQASDNKAVSLADIIVLAGCAGVE 557
Db 654 EGAR-SAEPSDSTQLAFIQALAKA--TLNDAADYFRD-----ILAG--NVE 695
Qy 558 KAAKDAGHEVQVFPNPGRADATAEOTDVEAFE---ALEPAD---GFRNYIKPEHKVSA 610
Db 696 -----GLTVDPDLRWALTIARGDIEAEDAIAELSRDSSSASFLASLRAGAAVNT 749
Qy 611 EEMLVDRQAQLLSAPENTALVGOM-----RVLTGNTYDGSOR-----GV 649
Db 750 EE-----VKAAYKHVTAVDGSLNLELRHIEGLTFTGSSSELLOAYNEQYFEILDDV 802
Qy 650 FTKPKQSLNDFVNLDLNTKWRASDE----SDKVPEG 684
Db 803 WANFSGEMAQOI---VLGLFPPSNVSEGLKRTDFLDG 838

RESULT 8
US-10-128-714-8481
; Sequence 8481, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroskin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
```



```

: FILE REFERENCE: 10182-018-999
: CURRENT APPLICATION NUMBER: US/10/128,714
: CURRENT FILING DATE: 2002-04-23
: PRIOR APPLICATION NUMBER: US 60/285,697
: PRIOR FILING DATE: 2001-04-23
: PRIOR APPLICATION NUMBER: US 60/287,066
: PRIOR FILING DATE: 2001-04-27
: PRIOR APPLICATION NUMBER: US 60/295,890
: PRIOR FILING DATE: 2001-06-05
: PRIOR APPLICATION NUMBER: US 60/303,899
: PRIOR FILING DATE: 2001-07-09
: PRIOR APPLICATION NUMBER: US 60/316,362
: PRIOR FILING DATE: 2001-08-31
: NUMBER OF SEQ ID NOS: 8603
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 8481
: LENGTH: 976
: TYPE: PRT
: ORGANISM: Aspergillus fumigatus
: US-10-128-714-8481

Query Match      3.0%; Score 121; DB 15; Length 976;
Best Local Similarity 19.2%; Pred. No. 0.12;
Matches 160; Conservative 99; Mismatches 255; Indels 318; Gaps 45

QY      130 REAPLN-----SHPDNANLDKARLLLPKQYGRKTSWADMLLTGNVALETWGF 180
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      7 REFYNPQAINALSFSPHPSAEL-----ACRGVPTTLRLAT----- 41

QY      181 KTFGAGGADVWPEEDVYGAETWLGDKRYEGDRELENPGLGAVQMGLIYV--NPR-G 237
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      42  --GRANDIEIWPMPGA-WFQTVVRGKK---DRSIE-----GLAWTLDDPPENG 85

QY      238 PNGKPPD-----IAAADIRETFGRMANDEETVALIAGGHTFGKTH 279
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      86 PDGTKIPGKLRFSIGYSTAVTEWNLEGRPMRHSSGNYG---EIWCIAAQPRWQATK 140

QY      280 GAADAE---KYVGREPAAGIEEMLGKWKTYG-----TG 311
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      141 RKDGKPLPAEGEYTCQH--LAACAGDSIVILSTAGDLKFLMRPSTRKARVLSVTF 199

QY      312 HGADITSGLEGAWTKTPTOMSNFFENLFGYEWELTKSPAGAYQ---WRPK----- 360
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      200 QNRNTIAGVADSSIRFLDRESSGQLRTI---SLKGPAGGKELLVWSVKLPDGTI 254

QY      361  ---DCAGAGTIPDAHDS--KSHAPMLTTDLALRMDPD-----YEKISRRYYENPDE 408
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      255 VSGDSAGEIRFWDAKNTSLIQRIQGHILADTLDAVSAAGDTVWSSGADORTVYRKSPGE 314

QY      409 FADAFKAWKTLTHR-----DMGPKV---RYLG----- 433
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      315 KGEKKGR-WAEVMHRRYTHDVKTFAYVETRDISIVVSGGPDASPVVLPREFGEKHHRK 373

QY      434  ---PEYPQ-----EDLIWDPIPDVSH--PLVDENDIEGLKAKIL----- 468
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      374 LSSLPQIPQLASSPSSRLVMSFWDSYIWR-----VSRGPISPHDNLEGGKHLRVGRVL 428

QY      469  ---ESGLTVSEL-----VSTAWASASTFRNSDKRG-----GANGAR 501
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      429 IQGEENTISAMLSADGKILIVATVSVKVFVSRBRKSDERGTLRTIKLDVPAAFSKDGR 488

QY      502 -IRLAPQKDW-----EYNNPQQLARVL-----KTLRGI 528
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      489 AVSISPSRWLCVVRPDSAIYLRVMPASSQDPKQIILPOLVKLDRAPTRHTRHKEASHGS 548

QY      529 QEDFNQA-----QSDNKAVSLADLIVLAGC--AGVEKAARDAGCHEVQVPPENPGRADATA 580
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      549 LGDYERTVRCVVFSENSRIILATGD---LSGCIDTLLKDVKDKSSAVASKNK--GAADSD 604

QY      581 EQDTVEAFEALPPAAG----FRNYIKPEHKVSAEMLVD-----RAQLLSLAPEMT 629
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      605 ESSD---DEDERPVYIDGERQYANADSP IPRLKSGLVLLSPRQSPAEKLLTNGADQSS 661

```

[illegible][illegible]

Db 429 -----RADAFYALYDRHITVDFAHPEADLSAYPLVVPALYLMTEAAGNNLEEVYENG 483
QY 428 K--VRYLGPVQEDLIWQDPIVDVSHPLVDNDIEGLKA-----KILESLTVS 475
Db 484 TLVVSFSGIVDEHDVHDGPPGAL-----RDVLGLTVVEFSPLLKDDRIYVTPGDS 537
QY 476 ELVSTAWA-----SASTFRNSDKRGANGARIRLAPQKDWVNNPQQLARVLTLE 526
Db 538 ELTGVMSEFVYPRGAETVWTYADGLTAGHPAVTRHRRFGESAWYVS-----TRLTAQ 590
QY 527 GQEDFNQASDNKAVSLADL 547
Db 591 GLDALLGRAANDARIAPRADL 611

RESULT 10
US-10-259-165-16
; Sequence 16, Application US/10259165
; Publication No. US20030135888A1
; GENERAL INFORMATION:
; APPLICANT: Zhu, Tong
; APPLICANT: Wang, Xun
; APPLICANT: Chang, Hur-song
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING
; FILE REFERENCE: 70030-NP
; CURRENT APPLICATION NUMBER: US/10/259,165
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/368,327
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 782
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 16
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-259-165-16

Query Match 2.8%; Score 113; DB 12; Length 331;
Best Local Similarity 20.4%; Pred. No. 0.11;
Matches 55; Conservative 26; Mismatches 97; Indels 92; Gaps 7;
QY 161 ISWADMLTGNVALETGKFTFGAGGRADVWEPEEDVYWGAEETWLGDKRYEGDRELE 220
Db 123 VSCADVLAFAARDALALVGNAYQVPGRRD----- 153
QY 221 NPLGAVQMGLIYVNPBPGKDPDIAAARDIRETFGRMANDEETVALLIAGGHTFGKTHG 280
Db 154 -----GNVSVAQETNGNLPPPSANVAQLNQMGKGLTQAEYVAL-SGAHTTIGVSHC 204
QY 281 AADAERYVGREPAAAGIEEMSLGWNKYTGTHGADTITSGLEGAWTKTPTQNSNFFENL 340
Db 205 SFSNRLYSSGPN-----GQDPSMDPSYVAALTTQCPOQ----- 239
QY 341 FGWEELTKSPAGYQWKPKDGAGACTIP-DAHDPSKSHAPF-----MLTTDLALR 390
Db 240 -----GQGP-----AAGWVPMDAVTPNAEDTNYAALVANRGLVSSDQALL 280
QY 391 MDPDYEKISRRYYENPDEFADAFAKAWYKL 420

Db 281 ADQTTAAQVGVYTNPNPDSFQTDFAAAWVKM 310
RESULT 11
US-10-259-165-360
; Sequence 360, Application US/10259165
; Publication No. US20030135888A1
; GENERAL INFORMATION:
; APPLICANT: Zhu, Tong
; APPLICANT: Wang, Xun
; APPLICANT: Chang, Hur-song
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING
; FILE REFERENCE: 70030-NP
; CURRENT APPLICATION NUMBER: US/10/259,165
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/368,327
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 782
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 360
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-259-165-360
Query Match 2.8%; Score 113; DB 12; Length 331;
Best Local Similarity 20.4%; Pred. No. 0.11;
Matches 55; Conservative 26; Mismatches 97; Indels 92; Gaps 7;
QY 161 ISWADMLTGNVALETGKFTFGAGGRADVWEPEEDVYWGAEETWLGDKRYEGDRELE 220
Db 123 VSCADVLAFAARDALALVGNAYQVPGRRD----- 153
QY 221 NPLGAVQMGLIYVNPBPGKDPDIAAARDIRETFGRMANDEETVALLIAGGHTFGKTHG 280
Db 154 -----GNVSVAQETNGNLPPPSANVAQLNQMGKGLTQAEYVAL-SGAHTTIGVSHC 204
QY 281 AADAERYVGREPAAAGIEEMSLGWNKYTGTHGADTITSGLEGAWTKTPTQNSNFFENL 340
Db 205 SFSNRLYSSGPN-----GQDPSMDPSYVAALTTQCPOQ----- 239
QY 341 FGWEELTKSPAGYQWKPKDGAGACTIP-DAHDPSKSHAPF-----MLTTDLALR 390
Db 240 -----GQGP-----AAGWVPMDAVTPNAEDTNYAALVANRGLVSSDQALL 280
QY 391 MDPDYEKISRRYYENPDEFADAFAKAWYKL 420
Db 281 ADQTTAAQVGVYTNPNPDSFQTDFAAAWVKM 310

RESULT 12
US-10-156-761-9790
; Sequence 9790, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI

```

; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9790
; LENGTH: 1104
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9790

Query Match          2.8%; Score 112.5; DB 15; Length 1104;
Best Local Similarity 18.7%; Pred. No. 0.85;
Matches 136; Conservative 68; Mismatches 256; Indels 267; Gaps 31;

QY 99 HYGPFIRMAHWSAGTYR-----ICDGRGG-----GGSGSQRFAPLNSW 137
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 232 HFTAREDR-AFKSTGTWQDRNLNPGSTASGGTGGSGQGRPVAGGAGGYEFAPGSG- 289
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 138 PDNANLDKARLLMPKIKYGRKISWADMLILTNVALETMGKTF----- 183
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 290 -----PVNVKVG--ISYVSQAGAEANLAVENPRGRSPFAVRAARRAWRER 333
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 184 ----GFAGGRADWEEDEYVWGAETEWL-----GKRYEGDRELENPLG---AVQM 228
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 334 LGAVGVGGTAE-----DRTFTYALYHALLHPNVISDADRRYRGSDKRVHVYGRHRAQY 389
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 229 G-----LIYVNP-E-GPNGKDPDIAARD-----IRETFGRMAMDEET 265
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 390 GTFSGHDYRSQVQLTLDPRTGSDIAQSLLELARONGGINDRWLHGASGTHVMGDPS 449
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 266 VALIAGGHTFGKT-----HGAADA-----EKYV----- 288
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 450 PAALAGTRAFGGTDFDLRGALDSLVAATVPTSQDLSAAGKPVLSVGQRPQLDKYLNRY 509
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 289 -----GREPAAGIEEMSLGKNTYG----- 309
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 510 MPSVSNAGGAETLEMSGADFALSQARAAGEKETATFARRSQWQNNFNIAAAPSPS 569
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 310 --TGHGADTITSG-----LEGATK--TPTQSNFFENLFG-YEWELTKSPAGAYQ-- 356
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 570 ARVSGSGYLNSGYIANRRADGSMVTGFTPAT-GNGFVEGTAQAQYTWVQHNPAGLEAAM 628
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 357 -----WKPDKGAGACT-----IPDAHDP---SKSHAPFMLTDLAL 389
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 629 GGRDAALDRLDAPFHDSDGSWAFTGGGDKSELDNEPSINVPYLYAYAGAPYKTQETVRA 688
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 390 RMDPDYKISRYYENDEPADAFAKAWYKLTDRMGPKVRYLGPVPEQEDLIWQDPI-P 448
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 689 AMDKLSTRPGGTPGNDLGA---MSSWYFSALGMYPOV-----PSRAELVLASPLFP 739
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 449 DVSHPLVDENDIEGLAKILESLTVSEL-----VSTAWASASTERNKDRKGANGARI 502
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 740 RVEIRPHGNDIEIRAAGAAKDAPYVRSLSKVDGRTSDRSWLPASFYRD---GGRLDYTL 795
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 503 RLAPQDKMEVN-----NPOOL-----ARVLKTLGIGQE 530
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 796 SATPNRAWGADAADAPPFRAGBQPYQIGVPTTATVAPGGSTKIDIRALALSGGTGPEV 855
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 531 DFNQASDNKAVSLADLIVLAGCAGYKAAKADGAGHVQVFPNPRGRADATAEQDZFAEA 590
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 856 RFRVETPAGVTATPAEGTSDGAQQITVTAADAARQ---GFYDVKVTVTSPDTSYEQPVA 912
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 591 LEPAADG 597
   | | | | |
Db 913 LTVAAFG 919
```

```

RESULT 13
US-10-156-761-9552
; Sequence 9552, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAVOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9552
; LENGTH: 728
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9552

Query Match          2.8%; Score 111; DB 15; Length 728;
Best Local Similarity 20.2%; Pred. No. 0.6;
Matches 166; Conservative 82; Mismatches 252; Indels 320; Gaps 48;

QY 5 KHSGSFTYNTGCKCPF--TGSLSKOSAGGTT-----KNRDWPNMNLGILRQHSLS 57
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 47 RHRWQLTFT-AGRDTFQVSGRTGRVTVTGGTPTATQLTGLNY-----LR 91
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 58 DPNDPFDVAEFPKLDLAAVKDKDLAALMT-----DSQMWPAHYGHYGPFI 105
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 92 NIADADINWAG--RQLRLPRALPLAGTVTRRANVPHRFALNDTND-----GYTGPYH- 142
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 106 RMAWHSAGTYRIGDGRGGSGSQRFAPLNSWPDNANLDKARLLMPKIKYGRKISWAD 165
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 143 --DW-----TY-----W--EREGLVIAL-----HG-----YNE 161
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 166 LMLITGNVALETMGKTFGAGGRADWV--EPBEDVTWGAEE-----TEWLGKRYE 214
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 162 VLVTQCADALHRRVFQFGYTGDELRKWIPGPAHQPWLLQNLNSAFPDPVSQQLDARAA 221
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 215 GDRELENPLGAVOMGLIYVNPGE--PNGKDPDIAAARDI-----RETF 255
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 222 LGRRIANRLRELGMTVPFGYFGTVPPGFADRNAGAHVTPQGTWGMGFARPDWLDPRTEHF 281
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 256 GRNA-----MNDRETVALIAGGHTFGKT-----HGAADAKEYVGREPAAAGIEE- 299
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 282 TRVAAAFYRIQDE---MFGGASTRYKMDLLHEGGSGDVPVG--DAAKGVERALRAHP 335
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 300 ----MSLGWKN-----TYGTGHGADTITSL-----EGAWTKTPTQSNFFEN 339
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 336 GAWVVLGWQHNPPRAIVDAVDKDRMLVVDGLCDRPFKVTDRADWHGTP-----Y 386
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 340 LFGEYENEL-----TKSPAGAYQ-WKPKDGA---GAGTIPDAHDPKSHAPFMLTDT 386
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 387 AFGSIWFGGHTTLCGANTPDWASLYERWTRPGSTLRGVALLEPAAD--NNPAFAALFSE 444
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 387 LALRMDPDYKISRYYENDEPADAFAKAWYKLTDRMGPKVRYLGPVPEQEDLIWQDPI 446
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 445 LAWRE-----GDLDRAWFAWAHS-----RYGGRD--PHEAAW--- 477
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 447 IPDVSHPLVDENDIEGLKAKILESLTVSELVSTAWASASTERNKDRKGANGARIAP 506
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 478 -----DI-----LRTAYGTTTRADSWSEGADGLFARPSLAA 509
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 507 QK--DWEVN----NPOQ-----LARVLKTLGIGQ-----EDFNQASDNKAVSL-- 544
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Db 510 TKAASWPKRLRYRPEFEPPALGELLKVRPGRLGSSAYRRDLLDVAQOALSNNRRSVLLPQ 569
Qy 545 -----ADLVLAGCAGVEKA-----AKDAGHEVQVFPNPGRA-DATAPQTDVEA 587
Db 570 IRTAYEAKDTAREDLRTGVLALMDLLEALLATDSRHLGLRWVADARAWGASAAERDLA 629
Qy 588 FEAL-----EPAAD-GFRNYTKPEHKVSAEMLVDRAQLLSLSAPEMTALVGM-RV 637
Db 630 YDALSLTLTVMGTACADAGLRDY-----ANRENAGLVGGLYRL 667
Qy 638 LGTNY-----DGSQHGVFTNKPGLSNDFFVNLDDLNTKW 672
Db 668 RWSTYFAELRSASREGTRPK-----TDWFA-----LEDWR 698

RESULT 14

US-10-156-761-11828
; Sequence 11828, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 11828
; LENGTH: 929
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-11828

Query Match 2.88; Score 110.5; DB 15; Length 929;

Best Local Similarity 19.18; Pred. No. 0.98; Mismatches 359; Indels 247; Gaps 36;
Matches 157; Conservative 103;

Qy 33 GGTNRDRWPNMLNLGILRHSLSLSD-----PNDPDFDYAEFFKLDLAUVKKDLAAL 85
Db 10 GGTGNQPPFGAQCACPRNSRSMSEHRKLPQPEGRAAARGQSQAPSGRRAPRAAT 69
Qy 86 MTDSDQDWWPAD-----YGHYGPFFIRMAWHS--GTYRIGD--GRGGGSGSQRFAPLNSW 137
Db 70 GSPSDSYGADGEERYPG--GRAEARASQRPERRAADAGRGSGGGGRRGP--GG 125
Qy 138 PDNAN-----LDKARLLWPVKQYGRK-----TSWADMLTLTGNVA 174
Db 126 PGGNPGRGGRGSRVPRKPFIDYPRAGKYGAARWLPWRVLVTGLCTAFTGSMVAVAGVA 185
Qy 175 LETMGFTFGAGGRADWPEEDVYMGATEMLGDKRYEGDRENLPLGAVQVGLIY-- 232
Db 186 YALGVVPKVALTA-----KAQNNVYWKDSQWVATGETNRQIVN-ISOIPKAMYAV 238
Qy 233 VNPGEPPNGKPPDIAAANDIRETGRMAWDEETVALIAGHTFGKTHGAADAKEYG--- 289
Db 239 ISOENKTFETDSGVDPKGIARAFNMA-----RGGETQG---GSTITQYVKNAM 285
Qy 290 -----REPAAGTEMSLGHKNKTYGTGHGADTTISGL-----PGAW 325
Db 286 LDDQSQTISRKKEIFVAIVKGVAKVPKDEIMAGYLSNAYYGRGAYGQAGARAYFDRDAK 345
Qy 326 TKPTQWNSNFFENLFGYEWELTKSPAGAYQWPKDGAGAGATGTPDAHDPSKSHAPFLMT 385

Db 346 NLNFGQCA-----FLAAMLKGATYYDP-----AGATSLDPAATSEANSRAKLQM 390
Qy 386 DLALRMDPDY---EKISRYY-----ENPDEFADAFAKAKYKLTHRDMGPKVRYLGPV 436
Db 391 QDTLDKEVEYCHLDEATRNKYTELKPVENPRSNARLSQVGYLV---DLAKAYLVNNSI 447
Qy 437 PQEDL-----IW-----ODPIPOVSHPLVDENDIEGLKAKILESGLVSELYS 479
Db 448 SADDLRGGYSTYTTFDKKKVTELEDAVKAVRKNKIKPKQRPDQDTYVQFGGASVDPTTG 507
Qy 480 TAW-----ASASTFRNSDRGGANGARIR---LAPQKDWENNPPOOLARVLKTLGIIQ 529
Db 508 AIKAIYGGEDATKHTNNADQGTGAQVGSFTFFVLAAMSGKKDPDL----- 555
Qy 530 EDFNQAOQSDNKAVSLADLIVLAGCAGVEKAADAGHEVQVFPNPGRADATAEQTDVEAFE 589
Db 556 -DSSQSDERTVVS-----PKSLYSGKNKLIKIEDYKGNV-----WTDNKGKE 596
Qy 590 ALEPAADGFRNYIKPEHKVSAEML-----VDRQAQLLSLA----- 625
Db 597 WLO--TNDGGSYNAPSYKIDLEAMRVSAANSFAVLQGLMDVGLDKVRDAALSAGILKTSLA 655
Qy 626 -----PEMTALVGMRLVG-----TNYDGS--OHG-----V 649
Db 656 SANTPFSIGTSDPSAIRMAGAYATFAASGKQRPDPYSVEKVTSKDGPVQHSIDIAKYDDA 715
Qy 650 FTNKPQLSNDFFVNLDDLNTKKRASDESCKVFEGRDF--KTGEVKWSGTREVLDIFGSNS 707
Db 716 FTAKVADNVTDLTKTVVDQGTGTAQ-----LTGREVAGKTGTT--DGNSANFVGYP 767
Qy 708 ELRALAEVYGCADSEEFKRVDFKAW 733
Db 768 QLSTSTMYRMDNESSKKKROFLEMY 793

RESULT 15

US-10-156-761-13251
; Sequence 13251, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13251
; LENGTH: 1208
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13251

Query Match 2.78; Score 110; DB 15; Length 1208;
Best Local Similarity 21.58; Pred. No. 1.7;
Matches 94; Conservative 41; Mismatches 161; Indels 142; Gaps 17;

Qy 6 HSGSSTYNTWTGKCPF--TGGSLKQSGAGGTGKRDWPNMLNLGILRHSLSLSDPNDPD 63
Db 542 HGTAATVTTGDSKAPVGETSGQIVASAGAPV-----AHTA----- 578
Qy 64 FDVAEEFKKLDLAUVKKDLAALMTDS-----ODWNPADYGHYGPFFIRMAWHSAG 113
Db 579 FGLVKEERYSLTVHVDRAAGATAAGLVVQRLAAGTDPYPASVGDSTLALRLA---PG 635

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 7, 2003, 19:39:02 ; Search time 29.3425 Seconds
(without alignments)
2441.707 Million cell updates/sec

Title: US-09-884-889-8
Perfect score: 4002
Sequence: 1 MENTHSCSYNTNTGGKC.....VKDFVKAQKVDLDRFDLK 745

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2787.5	69.7	727	2 F95422	catalase (EC 1.11.1)
2	2759.5	69.0	735	2 B83763	catalase BH0906 [i
3	2714.5	67.8	735	1 J50520	catalase (EC 1.11.1)
4	2714.5	67.8	751	4 T43873	catalase (EC 1.11.1)
5	2714.5	67.8	751	4 T43874	catalase (EC 1.11.1)
6	2714.5	67.8	751	4 T43878	catalase (EC 1.11.1)
7	2714.5	67.8	751	4 T43880	catalase (EC 1.11.1)
8	2714.5	67.8	751	4 T43881	catalase (EC 1.11.1)
9	2714.5	67.8	752	4 T43877	catalase (EC 1.11.1)
10	2714.5	67.8	753	4 T43882	catalase (EC 1.11.1)
11	2706.5	67.6	781	2 F82584	catalase/peroxidase
12	2691	67.2	737	2 A87626	catalase (EC 1.11.1)
13	2544	63.6	740	1 A40662	catalase (EC 1.11.1)
14	2541.5	63.5	746	1 A47685	catalase (EC 1.11.1)
15	2444.5	61.1	726	2 A10936	catalase (hydroper
16	2439.5	61.0	740	2 T45091	catalase (EC 1.11.1)
17	2424.5	60.6	726	2 G91237	hydroperoxidase Hp
18	2423.5	60.6	726	1 C5ECPH	catalase (EC 1.11.1)
19	2421.5	60.5	731	2 T44846	catalase (EC 1.11.1)
20	2417.5	60.4	726	2 C85085	catalase, hydroper
21	2417	60.4	724	2 B82186	catalase/peroxidase
22	2407	60.1	727	1 C5EBHT	catalase (EC 1.11.1)
23	2401.5	60.0	720	2 T44562	catalase (EC 1.11.1)
24	2380.5	59.5	737	2 AC0403	catalase (EC 1.11.1)
25	2361.5	59.0	754	2 S75113	catalase (EC 1.11.1)
26	2312	57.8	736	2 T00313	catalase (EC 1.11.1)
27	2265	56.6	741	2 A69529	catalase (EC 1.11.1)
28	2233.5	55.8	720	2 S71130	catalase (EC 1.11.1)
29	2203.5	55.1	731	2 C98160	catalase (AB033631

ALIGNMENTS

RESULT 1

F95422

catalase (EC 1.11.1.6) [imported] - Sinorhizobium meliloti (strain 1021) magaplasmid
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 14-Sep-2001
C:Accession: F95422
R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; B
.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surryski, R.; Wells, D.H.; Yeh, K
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meli
A:Reference number: A95262; MUID:21396509; PMID:11481432
A:Accession: F95422
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-727 <NUR>
A:Cross-references: GB:AE006469; PIDN:AAK5944.1; PID:g14524460; GSPDB:GN00165
A:Experimental source: strain 1021, megaplasmid pSymA
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Anpe, F.; Barloy-Hubl
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: SWA2379
A:Genome: plasmid
C:Superfamily: catalase HPI
C:Keywords: oxidoreductase

Query Match 69.7% ; Score 2787.5; DB 2; Length 727;

Best Local Similarity 71.1% ; Pred. No. 5.3e-182;

Matches 521; Conservative 72; Mismatches 127; Indels 13; Gaps 6;

QY	14	TWTGCKPPTGSGKOSAGGGTKNRDWPNNLNLGTLROHSSISDPNDPDFDYAEFFKL 73
DB	5	SDSACKCP-----VAHTAPGRSNRDNWPDQJDVQLHRHSGLSPLGNTFTFAEKKL 59
QY	74	DIAAVKDLAALMTDSQDWWPADYGHYGFPIFIRMAHWSAGTYRIGDGRGGSGSORFAP 133
DB	60	DLDALKRLRALMTDSQDWWPADYGHYGLGFIRMAHWSAGTYRITDGRGAGGQQORFAP 119
QY	134	LNSWPDNANLDKARLLNPIKOKYGRKISWADMLITGNVALETMGFTGTFAGGRADYV 193
DB	120	LNSWPDNANLDKARLLNPIKOKYGRKISWADMLITGNVALETMGFTGTFAGGRADYV 179
QY	194	EPEEDVYGAETEWLGDREYEGDRELENPLGNVGLIYVNEPGNKGKPDPTAAARDIRE 253
DB	180	EPEE-LFWGPEGTWLGDERYSERQISEPLAAVQGLIYVNEPGNKGKPDPTAAARDIRE 238

F:101/Active site: His (distal axial ligand) #status predicted
F:264/Binding site: heme iron (His) (proximal axial ligand) #status predicted
F:315/Active site: Trp (tryptophyl radical intermediate) #status predicted

Query Match 67.8%; Score 2714.5; DB 1; Length 735;
Best Local Similarity 68.7%; Pred. No. 5.1e-177;
Matches 511; Conservative 75; Mismatches 139; Indels 19; Gaps 6;

QY	1	MENHHKSGSSTYNTWTGKCPPTGGSLKQSGAGGTNRDWPWMLNLGILRHSSLSDPN	60
DB	1	MENQ-----NRQNAACQPHGVSNTQSS-NRTTNKDWPNQLNLSLHQHDKRNTPH	51
QY	61	DPDFDYAEFEKLDLAALKVKKDLAALMTDSQDWWPADYGHYGFPIRMAMHSAGTYRIGDG	120
DB	52	DEEFNYAEFEKLDYALKEDLTKMTESQDWWPADYGHYGLPIRMAMHSAGTYRIGDG	111
QY	121	RGGGSGSQRFAPLNSWPDNANLDRLLWPIKQYGRKISWADLMILTGVALETMGF	180
DB	112	RGASTGTQRFAPLNSWPDNANLDRLLWPIKQYGRKISWADLFILAGNVAIESMGG	171
QY	181	KTFGAGGADYVWEPEEDYWGAEWGLDGRKRYEGDRELENPLGAVQMGLIYVNEP	240
DB	172	KTFGCGGRVYVHPEEDYWGSEKWLASERYSGDRELENPLAAVQMGLIYVNEP	231
QY	241	KDPDIAAADIRETFGRMANDEETVALTAGGTHGKTHGAADAKEYYVREPAAGIEM	300
DB	232	KDPKAAADIRETFGRMANDEETVALTAGGTHGKTHGAAGAPG-THVGEPEEAP	290
QY	301	SLGWNTYGTGHGADTITSGLEGAWTKPTQNSNFFENLFGYEWELTKSPAGAYQWKPK	360
DB	291	GLGWISSYKGGKSDTITSGLEGAWTKPTQNDTSYFDMFGYDWMWLTSPAGAWQWAV	350
QY	361	DGAGAGTTPDAHDPSKSHAPFMTDLALRMDPDYKISRRYYENPDEPADAFKAWYKL	420
DB	351	DPDEKDLAPDAEDPSKVPMTMTDLALRFDPEYEKIARRPHQNPPEEFAFAFAWEKL	410
QY	421	THRDMGPKVRYLGPVPEQEDLIWQDPIPDVSHPLVDENDIEGLKAKILESGLTYS	480
DB	411	THRDMGPKTRYLGPVPEKEDFIWQDPIPEVDYELT-EAEIEEIKAKILNSGLTV	469
QY	481	ANASASTFNSDKRGANGARILAPQKDWENYNNPQALRVLTLEGIEDFNQASDNK	540
DB	470	ANASASTFNSDKRGANGARILAPQKDWENYNNPQALRVLTLEGIEDFNQASDNK	523
QY	541	AVSLADLIVLAGCAGVERAAKADAGHEVQVFPNPGRADATAEOTDVEAFPAEADGFRN	600
DB	524	KVSIADLIVLGSAAVEKAARDAGFDVKVPFPFGRGDQTEQTDVESFAVLEPFDGFRN	583
QY	601	YIKPEHKVSAEMLVDRAQLLSAPENTALVGGMRVLTGTYDGSQHGVTNKPGLSND	660
DB	584	YQKQYVPPPEELLVDKALGLTAPMTLVGGLRVLGANTRDPLPHGVFTDRIGVLTND	643
QY	661	FFVNLDLNTRKRADESQVFEGRDFTKTEGVKWSCTRVLDLIFGNSSELRAEA	720
DB	644	FFVNLDMNYEWVPTDSG--IYEIRDRKTGEVYRWATRVLDLIFGNSILRS	701
QY	721	SEEKFKVDFKAWAKVMDLDRDL 744	
DB	702	NOEKFVRDFINAWKVMNADRFDL 725	

RESULT 4
T43873
catalase (EC 1.11.1.6) [imported] - synthetic
C:Species: synthetic
A:Note: cat gene engineered and expressed in Bacillus stearothermophilus
C:Date: 24-Mar-2000 #sequence_revision 24-Mar-2000 #text_change 23-Mar-2001
C:Accession: T43873
R:Matsuura, T.; Miyai, K.; Trakunaleaamsai, S.; Yomo, T.; Shima, Y.; Miki, S.; Yamamoto
Nat. Biotechnol. 17, 58-61, 1998
A:Title: Evolutionary molecular engineering by random elongation mutagenesis.
A:Reference number: 222644; MUID:99116998; PMID:9920270
A:Accession: T43873

A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-751 <MAT>
A:Cross-references: EMBL:AB020075; NID:g4140057; PIDN:BAA36987.1; PID:g4140058
C:Comment: For the wild-type sequence, see PIR:JS0520.
C:Genetics:
A:Gene: cat
C:Keywords: oxidoreductase

Query Match	67.8%;	Score	2714.5;	DB	4;	Length	751;
Best Local Similarity	68.7%;	Pred.	No. 5.3e-177;				
Matches	511;	Conservative	75;	Mismatches	139;	Indels	19;
Gaps	6;						
QY	1	MENHHKSGSSTYNTWTGKCPPTGGSLKQSGAGGTNRDWPWMLNLGILRHSSLSDPN	60				
DB	1	MENQ-----NRQNAACQPHGVSNTQSS-NRTTNKDWPNQLNLSLHQHDKRNTPH	51				
QY	61	DPDFDYAEFEKLDLAAYKDLAALMTDSQDWWPADYGHYGFPIRMAWHSAGTYRIGDG	120				
DB	52	DEEFNYAEFEKLDYALKEDLRKLTWESQDWWPADYGHYGLPIRMAWHSAGTYRIGDG	111				
QY	121	RGGGSGSQRFAPLNSWPDNANLDRALLWPIKQYGRKISWADLMILTGVALETMGF	180				
DB	112	RGASTGTQRFAPLNSWPDNANLDRALLWPIKKYGNKISWADLFILAGNVAIESMGG	171				
QY	181	KTFGAGGADYVWEPEEDYWGAEWGLDGRKRYEGDRELENPLGAVQMGLIYVNEP	240				
DB	172	KTFGCGGRVYVHPEEDYWGSEKWLASEYSGDRELENPLAAVQMGLIYVNEP	231				
QY	241	KDPDIAAADIRETFGRMANDEETVALTAGGTHGKTHGAADAKEYYVREPAAGIEM	300				
DB	232	KDPKAAADIRETFGRMANDEETVALTAGGTHGKTHGAAGAPG-THVGEPEEAP	290				
QY	301	SLGWNTYGTGHGADTITSGLEGAWTKPTQWNSNFFENLFGYEWELTKSPAGAYQWKPK	360				
DB	291	GLGWISSYKGGKSDTITSGLEGAWTKPTQMDTSYFDMFGYDWMWLTKSPAGAWQWAV	350				
QY	361	DGAGAGTTPDAHDPSKSHAPFMTDLALRMDPDYKISRRYYENPDEPADAFKAWYKL	420				
DB	351	DPDEKDLAPDAEDPSKVPMTMTDLALRFDPEYEKIARRPHQNPPEEFAFAFAWEKL	410				
QY	421	THRDMGPKVRYLGPVPEQEDLIWQDPIPDVSHPLVDENDIEGLKAKILESGLTYS	480				
DB	411	THRDMGPKTRYLGPVPEKEDFIWQDPIPEVDYELT-EAEIEEIKAKILNSGLTYS	469				
QY	481	ANASASTFNSDKRGANGARILAPQKDWENYNNPQALRVLTLEGIEDFNQASDNK	540				
DB	470	ANASASTFNSDKRGANGARILAPQKDWENYNNPQALRVLTLEGIEDFNQASDNK	523				
QY	541	AVSLADLIVLAGCAGVEKAADAGHEVQVFPNPGRADATAEOTDVEAFPAEADGFRN	600				
DB	524	KVSIADLIVLGSAAVEKAARDAGFDVKVPFPFGRGDQTEQTDVESFAVLEPFDGFRN	583				
QY	601	YIKPEHKVSAEMLVDRAQLLSAPENTALVGGMRVLTGTYDGSQHGVTNKPGLSND	660				
DB	584	YQKQYVPPPEELLVDKALGLTAPMTLVGGLRVLGANTRDPLPHGVFTDRIGVLTND	643				
QY	661	FFVNLDLNTRKRADESQVFEGRDFTKTEGVKWSCTRVLDLIFGNSSELRAEA	720				
DB	644	FFVNLDMNYEWVPTDSG--IYEIRDRKTGEVYRWATRVLDLIFGNSILRSV	701				
QY	721	SEEKFKVDFKAWAKVMDLDRFDL 744					
DB	702	NOEKFVRDFINAWKVMNADRFDL 725					

RESULT 5
T43874
catalase (EC 1.11.1.6) [imported] - synthetic
C:Species: synthetic
A:Note: cat gene engineered and expressed in Bacillus stearothermophilus
C:Date: 24-Mar-2000 #sequence_revision 24-Mar-2000 #text_change 23-Mar-2001
C:Accession: T43874

R;Matsuura, T.; Miyai, K.; Trakulnaleamsai, S.; Yomo, T.; Shima, Y.; Miki, S.; YamamotoC Nat. Biotechnol. 17, 58-61, 1998
A;Title: Evolutionary molecular engineering by random elongation mutagenesis.
A;Reference number: Z22644; MUID:99116998; PMID:952070
A;Accession: T43874
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-751 <NAT>
A;Cross-references: EMBL:AB020079; NID:q4140065; PIDN:BA036991.1; PID:q4140066
C;Comment: For the wild-type sequence, see PIR:JS0520.
C;Genetics:
A;Gene: cat
C;Keywords: oxidoreductase

Query Match	67.8%;	Score	2714.5;	DB	4;	Length	751;	
Best Local Similarity	68.7%;	Pred. No.	5.3e-177;					
Matches	511;	Conservative	75;	Mismatches	139;	Indels	19;	
	Gaps	6;						
QY	1	MENKHSGS	TYNTGCGKCPFTGGSLKQSGAGGTTKRDWPNMLNLGILRHSSLS	DPN	60			
DB		1	MENQ-----NRQNAQC	PFHGSVTNQSS-NRTTKRDWPNQNLNLSILHQRDKTNPH	51			
QY	61	DPDFYABE	PKLDLAAVKKDLAALMTDSQDWWPADYGHYGPFFIRMAHSA	GTYRIGDG	120			
DB		52	DEEFNTAE	EFPQKLDYNAKEDLKLMTESQDWWPADYGHYGPLFIRMAHSA	GTYRIGDG	111		
QY	121	RGGGSGS	ORFAPLNSWPNANDKARLLIWPDKQYGRKISWADLMILGTNVA	ETWGF	180			
DB		112	RGASGTQ	RFAPLNSWPNANDKARLLWPKKYGKNSWADLFIAGNVAIES	MG	171		
QY	181	KTFGAGG	ADVYVEPDYVGAETWLGDKRYEGDRELENPICAVOMGLIYV	PEP	240			
DB		172	KTIFGGG	RVYVHPEEDYVWGSEKWLASERYSGDRELENPAAVOMGLIYV	PEP	231		
QY	241	KPDPIAA	ARDIRETFGRMANDEETVALIAGHTFGKTHGAADA	EYKVVGREPAA	300			
DB		232	KPDPKAA	ARDIRETFGRMGNDEETVALIAGHTFGKAHGAGPA-TVG	PEPEAA	290		
QY	301	SLGWKNT	YGTGHGADRTITSGEAGWTKTPTQWNSNPPENLFGYEWELTK	SPAGAYOKPK	360			
DB		291	GLGWISSY	GKGSDDITSGIEGAWPTPTQWNTSYFDMFGDWMUTK	SPAGAWQWAV	350		
QY	361	DGAGACT	IPDAHDPKSHAPFMLTTDLALRMDPDYKESIRRYENPDE	FADAFAKAYKL	420			
DB		351	DPDEKLA	PAEDPEKSKVPTMMVTTDLALRDFPEYKIRARFHQNP	EEFAEFAFAWEKL	410		
QY	421	THRDMG	PKVRYLGPEVPOEDLIWQDTPDYPHPLVDENDIEGLKAK	ILTESGLTVSELVST	480			
DB		411	THRDMG	PKTRYLGPEVPKEDFIWQDPIPEVDYELT-EAEIEEK	AKAKILNSGLTVSELVST	469		
QY	481	ANASAST	FRNSDKRGANGARIRLAPQKQWENVNPOOLARVLKTL	EGIQDFDQAO	SDNK	540		
DB		470	ANASAST	FRNSDKRGANGARIRLAPQKQWENVNEPERLAKVLS	YVEDIQREL-----PK	523		
QY	541	AVSLADL	IVLAGCAGYEKAADGACHEVOVPFNGRADATAEQD	TVFAFALEPA	ADGFRN	600		
DB		524	KVISADL	IVLGSSAAEKAARDAGFVPVFPFGRGDQEQTDV	ESFAVLEP	ADGFRN	583	
QY	601	YTKPEK	VSAEEMLVRAQLLSAPEMTALVGMRLVGTNYDSQ	HGVETNKGQ	LSND	660		
DB		584	YQKQES	VSPPEELLVDKQAQLLGITAPMTVLVGLRVLGANY	KDLPGVFTDRIGVLTND	643		
QY	661	FFVNL	LDLNTKWRASDESKVFEGRDKFTGVKWSGTRV	VDLIFGNS	SELRALAEVY	GCAD	720	
DB		644	FFVNL	DMNVEMVPTDSG--IVEIRDKKTGEVMTAT	RVDLIFGNS	SILASAEFFA	QADD	701
QY	721	SEKFKV	KDFKAWKAVMDLRDPL	744				
DB		702	NOEKFVR	DFINAWKYMNADRL	725			

RESULT 6
T43878

catalase [EC 1.11.1.6] [imported] - synthetic
C:Species: synthetic
A:Note: cat gene engineered and expressed in *Bacillus stearothermophilus*
C:Date: 24-Mar-2000 #sequence_revision 24-Mar-2000 #text_change_23-Mar-2001
C:Accession: T43878
R:Matsuura, T.; Miyai, K.; Trakunaleamsai, S.; Yomo, T.; Shima, Y.; Miki, S.; Yamam
Nat. Biotechnol. 17, 58-61, 1998
A:Title: Evolutionary molecular engineering by random elongation mutagenesis.
A:Reference number: Z26444; MUID:99116998; PMID:9920270
A:Accession: T43878
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-751 <MAT>
A:Cross-references: EMBL:AB020090; NID:g4140087; PIDN:BAA37002.1; PID:g4140088
C:Comment: For the wild-type sequence, see PIR:JS0520.
C:Genetics:
A:Gene: cat
C:Keywords: oxidoreductase

Query Match	67.8%	Score	2714.5	DB 4	Length	751
Best Local Similarity	68.7%	Pred.	No. 5.3e-177			
Matches	511	Conservative	75	Mismatches	139	Indels
					19	Gaps
QY	1	MENHKHSGSSYYNTYNTGKCPFTGGSILKQSAAGGCTKNRDWPNMNLNIGILIRQHSSLSDDPN	60			
Db						
Db	1	MENQ-----NRQNAACQCPFHGSVNTQSS--NRTTNKDWPNQLNLSTLHOHDKRTNPH	51			
QY	61	DPDFDYAEFEKFLDLAAVKKLALMTDSODWPADYGHYPPFFIRMAWHSAGTYRIGDS	120			
Db						
Db	52	DEEFNYAEFOKLDYWALKEDLRKLMTESQWPADYGHYGPLFIRMAWHSAGTYRIGDS	111			
QY	121	RGGGSGSQRTAPLNSWPNNDANLKDARLLLPKIQKYGRKISWADLMILTGNVALETMGF	180			
Db						
Db	112	RGGASTGTQRTAPLNSWPNNDANLKDARLLLPKIKKYGNKISWADLFILAGNVAIESMGG	171			
QY	181	KTFGAGGRADWPPEEDYVYGAEFWLGDKRDEGDELENPLGAVQMGLIYYNPEGPG	240			
Db						
Db	172	KTFGEGGGRVDWPEEDYVYGSKEWLASERYSGDRELENPLAAVQMGLIYYNPEGPDG	231			
QY	241	KPDPTAAARDIRETFGRMANDEETVALIAGGHTFGKTHGAADAERYVGREPAAAGIEEM	300			
Db						
Db	232	KDPKAAARDIRETFGRMANDEETVALIAGGHTFGKAGAGPA--THVGEPEAPAIIEAQ	290			
QY	301	SLGWNTYGTGHGADTTISGLEGAWTPTQWNNFFENLFGYEWELTKSPAGAYQWKPK	360			
Db						
Db	291	GLGMSSYSGKGGSDTTISGIEGAWTPTQWDTSYFDMLEFGYDMLTKSPAGAQWMAV	350			
QY	361	CGAGACTTPTDAHDSKSHAPMLTTDLALRMDPDYEKISRRYYENDEPDAFAFAKAYKL	420			
Db						
Db	351	DDEKDLAPDAEDPSKVPTMMMTTDLALRFPDEYEKIAARRFQNPDEEAEAFARAFWL	410			
QY	421	THRDMGPRVRYLGPVPEQDLIMQDPIPYVSHPLVDENDIEGKAKILESGTLVSELVST	480			
Db						
Db	411	THRDMGPKTRYLGPVPEKEDFIWQDPIPEVDYELT--EAEIEEIKALINSGTLVSELVKT	469			
QY	481	AWASASTFRNSDKRGGGANGARIRLAPQKDEVNANPQQLARVLKTLLEGIOEDFNQAQSDNK	540			
Db						
Db	470	AWASASTFRNSDKRGGGANGARIRLAPQKDEVNEPERLAKVLSVYEDIQREL-----PK	523			
QY	541	AVSLADLLVLACGACVEKAKDAGHEVQVFPNPGRADATAEQDTDVEAFALPEAPADGFRN	600			
Db						
Db	524	KVSIADLLVLGGSAAVEKKAARDAGFDYKVPFPFGRGDQEQDTDESFAVLPEFPADGFRN	583			
QY	601	YKPEHKYSAEMLVDRAQLLSLAPENTALVGMRVLGTNTYDGSOGHYVFTKPKQLSND	660			
Db						
Db	584	YQKQISYVPPBELVDRAQLLGUTAPENTVLVGLGRVLGANYRDLPHGVFTTDRIGLVLTND	643			
QY	661	FFVNLDLNTKWRASDESCKVFEGRDPKTCGEVKSQGT-RYVDLIFGSSNSELURALAAVYGAD	720			
Db						
Db	644	FFVNLDMNYEWVPTDSG--IYEIRDRKTCGEVWTATRVDLIFGSSNLSIRSYAEFYAODD	701			
QY	721	SEEFVKVDYKAWAKVMDLDRFDL	744			

Db 702 NOKEFVRDFINAWKVMNADREFDL 725

RESULT 7

T43880

catalase (EC 1.11.1.6) [Imported] - synthetic

C:Species: synthetic

A:Note: cat gene engineered and expressed in Bacillus stearothermophilus

C:Date: 24-Mar-2000 #sequence_revision 24-Mar-2000 #text_change 23-Mar-2001

C:Accession: T43880

R:Matsumura, T.; Miyai, K.; Trakulnaleaamsai, S.; Yomo, T.; Shima, Y.; Miki, S.; Yamamoto

Nat. Biotechnol. 17, 58-61, 1998

A:Title: Evolutionary molecular engineering by random elongation mutagenesis.

A:Reference number: Z22644; MUID:99116998; PMID:9920270

A:Accession: T43880

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-751 <MAT>

A:Cross-references: EMBL:AB020091; NID:94140089; PIDN:BAA37003.1; PID:94140090

C:Comment: For the wild-type sequence, see PIR:JS0520.

C:Genetics:

A:Gene: cat

C:Keywords: oxidoreductase

Query Match 67.8%; Score 2714.5; DB 4; Length 751;

Best Local Similarity 68.7%; Pred. No. 5.3e-177;

Matches 511; Conservative 75; Mismatches 139; Indels 19; Gaps 6;

QY 1 MENHKSSSTYNTNTGKCPFTGGSLKQSGAGGKTRDWPNNMLNLGILRQHSLSLSDPN 60

Db 1 MENQ-----NRQNAACCPHGSVTNQSS-NRTNKDWPNNQLNLSLHQRHDKRTPH 51

QY 61 DDPFYAEFFKDLAAVKKDLAALMTSDQWNPADYGHYGPFFIRMAWHSAGTYRIGDG 120

Db 52 DEEFYAEFFKDLAAVKKDLAALMTSDQWNPADYGHYGPFFIRMAWHSAGTYRIGDG 111

QY 121 RGGGSGSORFAPLNSPDNANLDRKLLPIKQYGRKISWADLMILTGNVALETWGF 180

Db 112 RGGASTGTQRFAPLNSPDNANLDRKLLPIKQYGRKISWADLMILTGNVALETWGF 171

QY 181 KTFGAGGRADYWEPEEDYVWGAETWLGDKRYEGDRELENPLGAVQMGILYVNPGE 240

Db 172 KTFGGGRVDVWHPPEEDYVWGAETWLGDKRYEGDRELENPLGAVQMGILYVNPGE 231

QY 241 KPDPYAAARDIRETGRMAMDEETVALIAGHTFGKTHGAADAKEYVGREPAAAGIEM 300

Db 232 KPDPYAAARDIRETGRMAMDEETVALIAGHTFGKTHGAADAKEYVGREPAAAGIEM 290

QY 301 SLGWKNTYGTGHGADTITSGLEGATKPTQSNFFENLFGYEWELTKSPAGAYQWPK 360

Db 291 GLGWISSYVKGKSGSTITSGLEGATKPTQSNFFENLFGYEWELTKSPAGAYQWPK 350

QY 361 DGAGAGTIPDAHDPKSHAPFMTTDLALRMDPDYKISRRYYENPDEFADAFKAWYKL 420

Db 351 DPDEKDLAPDAEDPSKVPYTMATTDALRMDPDYKISRRYYENPDEFADAFKAWYKL 410

QY 421 THRDMPKRYLGPVPEQEDLIWQDIPDVSHPLVDENDIEGLKAKILESGTVSELVST 480

Db 411 THRDMPKRYLGPVPEQEDLIWQDIPDVSHPLVDENDIEGLKAKILESGTVSELVST 469

QY 481 AWASASTFNSDKRGANGARIRLAPQKDWENYNNPQOLARVLKTLGEGEDFNQASDNK 540

Db 470 AWASASTFNSDKRGANGARIRLAPQKDWENYNNPQOLARVLKTLGEGEDFNQASDNK 523

QY 541 AVSLADLIVLAGCAGVEKAAKDAGHEVQVPPNPGRADATAEOTDVEAFALPAADGFRN 600

Db 524 KVSIALDLIVLGSAAVEKAAKDAGHEVQVPPNPGRADATAEOTDVEAFALPAADGFRN 583

QY 601 YIKPEHKVSAEEMLYDRAQLLSAPMTALYGGMRVLGTNTYDGSQHGVTYTKPQGLSND 660

Db 584 YKQEVSVPEELLVDKALGLITAPMTALYGGMRVLGTNTYDGSQHGVTYTKPQGLSND 643

QY 661 FFVNLDDLNTKWRASDESQVPEGRDFTGTGKVSNGTRVDLIFGNSLSLALAEVYGCAD 720

Db 644 FFVNLDDLNTKWRASDESQVPEGRDFTGTGKVSNGTRVDLIFGNSLSLALAEVYGCAD 701

QY 721 SEEKEVKDFVAKAWKVMOLDREFDL 744

Db 702 NOKEFVRDFINAWKVMNADREFDL 725

RESULT 8

T43881

catalase (EC 1.11.1.6) [Imported] - synthetic

C:Species: synthetic

A:Note: cat gene engineered and expressed in Bacillus stearothermophilus

C:Date: 24-Mar-2000 #sequence_revision 24-Mar-2000 #text_change 23-Mar-2001

C:Accession: T43881

R:Matsumura, T.; Miyai, K.; Trakulnaleaamsai, S.; Yomo, T.; Shima, Y.; Miki, S.; Yamam

Nat. Biotechnol. 17, 58-61, 1998

A:Title: Evolutionary molecular engineering by random elongation mutagenesis.

A:Reference number: Z22644; MUID:99116998; PMID:9920270

A:Accession: T43881

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-751 <MAT>

A:Cross-references: EMBL:AB020099; NID:94140105; PIDN:BAA37011.1; PID:94140106

C:Comment: For the wild-type sequence, see PIR:JS0520.

C:Genetics:

A:Gene: cat

C:Keywords: oxidoreductase

Query Match 67.8%; Score 2714.5; DB 4; Length 751;

Best Local Similarity 68.7%; Pred. No. 5.3e-177;

Matches 511; Conservative 75; Mismatches 139; Indels 19; Gaps 6;

QY 1 MENHKSSSTYNTNTGKCPFTGGSLKQSGAGGKTRDWPNNMLNLGILRQHSLSLSDPN 60

Db 1 MENQ-----NRQNAACCPHGSVTNQSS-NRTNKDWPNNQLNLSLHQRHDKRTPH 51

QY 61 DDPFYAEFFKDLAAVKKDLAALMTSDQWNPADYGHYGPFFIRMAWHSAGTYRIGDG 120

Db 52 DEEFYAEFFKDLAAVKKDLAALMTSDQWNPADYGHYGPFFIRMAWHSAGTYRIGDG 111

QY 121 RGGGSGSORFAPLNSPDNANLDRKLLPIKQYGRKISWADLMILTGNVALETWGF 180

Db 112 RGGASTGTQRFAPLNSPDNANLDRKLLPIKQYGRKISWADLMILTGNVALETWGF 171

QY 181 KTFGAGGRADYWEPEEDYVWGAETWLGDKRYEGDRELENPLGAVQMGILYVNPGE 240

Db 172 KTFGGGRVDVWHPPEEDYVWGAETWLGDKRYEGDRELENPLGAVQMGILYVNPGE 231

QY 241 KPDPYAAARDIRETGRMAMDEETVALIAGHTFGKTHGAADAKEYVGREPAAAGIEM 300

Db 232 KPDPYAAARDIRETGRMAMDEETVALIAGHTFGKTHGAADAKEYVGREPAAAGIEM 290

QY 301 SLGWKNTYGTGHGADTITSGLEGATKPTQSNFFENLFGYEWELTKSPAGAYQWPK 360

Db 291 GLGWISSYVKGKSGSTITSGLEGATKPTQSNFFENLFGYEWELTKSPAGAYQWPK 350

QY 361 DGAGAGTIPDAHDPKSHAPFMTTDLALRMDPDYKISRRYYENPDEFADAFKAWYKL 420

Db 351 DPDEKDLAPDAEDPSKVPYTMATTDALRMDPDYKISRRYYENPDEFADAFKAWYKL 410

QY 421 THRDMPKRYLGPVPEQEDLIWQDIPDVSHPLVDENDIEGLKAKILESGTVSELVST 480

Db 411 THRDMPKRYLGPVPEQEDLIWQDIPDVSHPLVDENDIEGLKAKILESGTVSELVST 469

QY 481 AWASASTFNSDKRGANGARIRLAPQKDWENYNNPQOLARVLKTLGEGEDFNQASDNK 540

Db 470 AWASASTFNSDKRGANGARIRLAPQKDWENYNNPQOLARVLKTLGEGEDFNQASDNK 523

QY 541 AVSLADLIVLAGCAGVEKAAKDAGHEVQVPPNPGRADATAEOTDVEAFALPAADGFRN 600

Db 524 KVSIALDLIVLGSAAVEKAAKDAGHEVQVPPNPGRADATAEOTDVEAFALPAADGFRN 583

QY 601 YIKPEHKVSAEMLVDRQAQLLSAPENTALVGGMRVLTGNYDGSQHGVTNKGQSLND 660
Db 584 YQOEYSVPPPELLVDRQAQLGLTAPENTVLVGLRVLGANYRDLPHGVFTDRIGVLTND 643
QY 661 FFVNLLDNTKWRASDESDESKVFEGRDFTKGTGVKWSGTRVDLIFGNSLSELRALAEVYGCAD 720
Db 644 FFVNLLDMNVEWPTDSG--IYEIRDRKTGEVWTATRVLDLIFGNSILRSAYAEFYAODD 701
QY 721 SEEFVKDFVKAKAKVMDLDRFDL 744
Db 702 NOEFVRDFINAWKVMNADRFDL 725

RESULT 9
T43877
catalase (EC 1.11.1.6) [imported] - synthetic
C:Species: synthetic
A:Note: cat gene engineered and expressed in *Bacillus stearothermophilus*
C:Date: 24-Mar-2000 #sequence_revision 24-Mar-2000 #text_change 23-Mar-2001
C:Accession: T43877
R:Matsumura, T.; Miyai, K.; Trakulnaleamsai, S.; Yomo, T.; Shima, Y.; Miki, S.; Yamamoto
Nat. Biotechnol. 17, 58-61, 1998
A:Title: Evolutionary molecular engineering by random elongation mutagenesis.
A:Reference number: 222644; MUID:99116998; PMID:9920270
A:Accession: T43877
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-752 <MAT>
A:Cross-references: EMBL:AB020089; NID:g4140085; PIDN:BAA37001.1; PID:g4140086
C:Comment: For the wild-type sequence, see PIR:JS0520.
C:Genetics:
A:Gene: cat
C:Keywords: oxidoreductase

Query Match 67.8%; Score 2714.5; DB 4; Length 752;
Best Local Similarity 68.7%; Pred. No. 5.3e-177;
Matches 511; Conservative 75; Mismatches 139; Indels 19; Gaps 6;
QY 1 MENHKSGSSTYNTGKCPFTGGSLKQSGAGGCTKRDWPNMLNGLIRQHSSLSDPN 60
Db 1 MEMO-----NRQNAACQPFHGSVTNQSS-NRTTNKDWPNQLNLSILHQRKTNPH 51
QY 61 DPDFDYAEFEKKLDLAALKDLMWDSQDWPADYGHYCPFFIRMAHWSAGTYRIGDG 120
Db 52 DEEFNYAEFEKKLDYALKEDLRKLMTESQDWPADYGHYCPFFIRMAHWSAGTYRIGDG 111
QY 121 RGGGSGSQRFAPLNSWPNANLDKARLLWPIKQYGRKISWADLMILTGNALETWGF 180
Db 112 RGGASTCTQRFAPLNSWPNANLDKARLLWPIKQYGRKISWADLMILTGNALETWGF 171
QY 181 KTFGAGGADVWEPEEDVYWGAEETWLGDKRYEGDRELENPLGAVQMGLIYVNPENG 240
Db 172 KTFGFGGRVVDVWHPEDVYWGSEKWLASERYSGDRELENPLAAVQMGLIYVNPENG 231
QY 241 KPPIAARDIRETFGRWANDDETVALIAGGHTFGTHGAADAETKVVGREPAAAGTEEM 300
Db 232 KPDPKAARDIRETFGRWANDDETVALIAGGHTFGTHGAADAETKVVGREPAAAGTEEM 290
QY 301 SLGKNTYGTGCHGADTTTSGEGAWTKTPTQWSNFFENFGYEWELTKSPAGAYQKPK 360
Db 291 GLGWISSYKAGKSDTTTSGEGAWTKTPTQWDSYFDMLFGVDWMLTKSPAGAWQWAV 350
QY 361 DGAGAGTIPDAHDPKSHAPFMTLTDLALRMDPDYKISRYYENPDFAFAKAWYKL 420
Db 351 DPDEKDLAPDAEDPSKKVPTMMMTDLALRMDPDYKISRYYENPDFAFAKAWYKL 410
QY 421 THRMGPKVYLGPEVPOEDLIWQDPIPDVSHPLVDENDIEGLKAKLTLESGLTVSELVST 480
Db 411 THRMGPKVYLGPEVPOEDLIWQDPIPDVSHPLVDENDIEGLKAKLTLESGLTVSELVST 469
QY 481 AWASATFRNSDRGGANGARIAPQKDWENVPNPOOLARVLKLTGLEGQEDFNOAQSDNK 540
Db 470 AWASATFRNSDRGGANGARIAPQKDWENVPNPOOLARVLKLTGLEGQEDFNOAQSDNK 523

QY 541 AVSLADLTVLGACGAVEKAARDAGHEVQVPPNPGRDATAEQTDVEAFEALEPAADGFRN 600
Db 524 KVSADLTVLGSSAAVEKAARDAGDFVKVPPFPGRGDQEQTDVESFAVLEPFDGFRN 583
QY 601 YIKPEHKVSAEMLVDRQAQLLSAPENTALVGGMRVLTGNYDGSQHGVTNKGQSLND 660
Db 584 YQOEYSVPPPELLVDRQAQLGLTAPENTVLVGLRVLGANYRDLPHGVFTDRIGVLTND 643
QY 661 FFVNLLDNTKWRASDESDESKVFEGRDFTKGTGVKWSGTRVDLIFGNSLSELRALAEVYGCAD 720
Db 644 FFVNLLDMNVEWPTDSG--IYEIRDRKTGEVWTATRVLDLIFGNSILRSAYAEFYAODD 701
QY 721 SEEFVKDFVKAKAKVMDLDRFDL 744
Db 702 NOEFVRDFINAWKVMNADRFDL 725

RESULT 10
T43882
catalase (EC 1.11.1.6) [imported] - synthetic
C:Species: synthetic
A:Note: cat gene engineered and expressed in *Bacillus stearothermophilus*
C:Date: 24-Mar-2000 #sequence_revision 24-Mar-2000 #text_change 23-Mar-2001
C:Accession: T43882
R:Matsumura, T.; Miyai, K.; Trakulnaleamsai, S.; Yomo, T.; Shima, Y.; Miki, S.; Yamamoto
Nat. Biotechnol. 17, 58-61, 1998
A:Title: Evolutionary molecular engineering by random elongation mutagenesis.
A:Reference number: 222644; MUID:99116998; PMID:9920270
A:Accession: T43882
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-753 <MAT>
A:Cross-references: EMBL:AB020119; NID:g4140145; PIDN:BAA37031.1; PID:g4140146
C:Comment: For the wild-type sequence, see PIR:JS0520.
C:Genetics:
A:Gene: cat
C:Keywords: oxidoreductase

Query Match 67.8%; Score 2714.5; DB 4; Length 753;
Best Local Similarity 68.7%; Pred. No. 5.3e-177;
Matches 511; Conservative 75; Mismatches 139; Indels 19; Gaps 6;
QY 1 MENHKSGSSTYNTGKCPFTGGSLKQSGAGGCTKRDWPNMLNGLIRQHSSLSDPN 60
Db 1 MEMO-----NRQNAACQPFHGSVTNQSS-NRTTNKDWPNQLNLSILHQRKTNPH 51
QY 61 DPDFDYAEFEKKLDLAALKDLMWDSQDWPADYGHYCPFFIRMAHWSAGTYRIGDG 120
Db 52 DEEFNYAEFEKKLDYALKEDLRKLMTESQDWPADYGHYCPFFIRMAHWSAGTYRIGDG 111
QY 121 RGGGSGSQRFAPLNSWPNANLDKARLLWPIKQYGRKISWADLMILTGNALETWGF 180
Db 112 RGGASTCTQRFAPLNSWPNANLDKARLLWPIKQYGRKISWADLMILTGNALETWGF 171
QY 181 KTFGAGGADVWEPEEDVYWGAEETWLGDKRYEGDRELENPLGAVQMGLIYVNPENG 240
Db 172 KTFGFGGRVVDVWHPEDVYWGSEKWLASERYSGDRELENPLAAVQMGLIYVNPENG 231
QY 241 KPPIAARDIRETFGRWANDDETVALIAGGHTFGTHGAADAETKVVGREPAAAGTEEM 300
Db 232 KPDPKAARDIRETFGRWANDDETVALIAGGHTFGTHGAADAETKVVGREPAAAGTEEM 290
QY 301 SLGKNTYGTGCHGADTTTSGEGAWTKTPTQWSNFFENFGYEWELTKSPAGAYQKPK 360
Db 291 GLGWISSYKAGKSDTTTSGEGAWTKTPTQWDSYFDMLFGVDWMLTKSPAGAWQWAV 350
QY 361 DGAGAGTIPDAHDPKSHAPFMTLTDLALRMDPDYKISRYYENPDFAFAKAWYKL 420
Db 351 DPDEKDLAPDAEDPSKKVPTMMMTDLALRMDPDYKISRYYENPDFAFAKAWYKL 410
QY 421 THRMGPKVYLGPEVPOEDLIWQDPIPDVSHPLVDENDIEGLKAKLTLESGLTVSELVST 480

Db 411 THRMGPKTRYLGPVPKEDFIWQDPIPEVDYELT-EAEIEIKAKILNSGLTSELVYKT 469
QY 481 AWASASTRNSDKRGANGARIRAPQKDWENNPQOLARVLKTELGTOEDFNOAQSDNK 540
Db 470 AWASASTRNSDKRGANGARIRAPQKDWENNPQOLARVLKTELGTOEDFNOAQSDNK 540
QY 541 AVSLADLIVLAGCAGVEKAAKADGHEVOVPPNPGRADATAEOTDVEAEPAADGPRN 600
Db 524 KVSADLIVLGGSAKAAKADGHEVOVPPNPGRADATAEOTDVEAEPAADGPRN 583
QY 601 YIKPEHKVSAEMLVDRQAQLLSAPEMTALVGMNRVLGTNDGSHQGVFTNKPGQLSND 660
Db 584 YKQEYSVPPRELLVDRQAQLLSAPEMTALVGMNRVLGTNDGSHQGVFTNKPGQLSND 660
QY 661 FVNLILLNTWRASDESKVFEGRDFTKGVKSGTRVDFLFGSNLSRLALAEVYGCAD 720
Db 644 FVNLILLNTWRASDESKVFEGRDFTKGVKSGTRVDFLFGSNLSRLALAEVYGCAD 720
QY 721 SEERFVKDFVAKAVMDLDRFDL 744
Db 702 NQERFVRDFINAWKVMNDRFDL 725

RESULT 11
F82584
catalase/peroxidase XF2232 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 24-Aug-2001
C:Accession: F82584
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A62515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: F82584
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-781 <SIN>
A:Cross-references: GB:AE004035; GB:AE003849; NID:9107371; PIDN:AAF85031.1; GSPDB:GN001
A:Experimental source: Strain 9a5c
R:Simpton, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H
as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, B
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovsky-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF2232
C:Superfamily: catalase HPI
C:Keywords: heme; iron; metalloprotein
F:126/Active site: His (distal axial ligand) #status predicted
F:313/Binding site: heme iron (His) (proximal axial ligand) #status predicted
F:364/Active site: Trp (tryptophyl radical intermediate) #status predicted

Query Match 67.6%; Score 2706.5; DB 2; Length 781;
Best Local Similarity 67.1%; Pred. No. 2e-176;
Matches 510; Conservative 86; Mismatches 131; Indels 33; Gaps 7;

QY 9 SSTVNTNGGCPFTGGSLSKOSAGGCTKNRQWPNMLNLGILROHSSLSDDNDPFDYAE 68
Db 30 ASSPETTSVAVKCP-----NKTAVEGTHNKQWPNQLRVDLHROHSSNSKLGETFDYAK 84
QY 69 EFFKLDLAAKRLDHALMTDSQDWWPADYHYGPPFFIRMAWHSAGTYRIGDGGGGSGS 128
Db 85 EFQKLDLAAKRLDHALMTDSQDWWPADYHYGPPFFIRMAWHSAGTYRIGDGGGGSGS 144

QY 129 QRAPLNSWPONANLDKARLLLPKQYGRKISWADLMILTGNVVALETMGFTFGFAGG 188
Db 145 QRAPLNSWPONVSLDKARRLLPILKKYGOQISWADLIVLSGNVALESMTGFTFGFAGG 204
QY 189 RADYWEPEEDYWGAEETWL-GDKRY-----EG-----DRELENPLG 224
Db 205 RVDYWEPEEDYWGAEETWLGGDVRVYGAIVSEGVHHPDEHGRGAKEAKANSRVLLENPLA 264
QY 225 AVQGLIYVNPPEGNGKPDPTAAARDIRETFORMAMNDEETVALLAGHTTGFCTGAADA 284
Db 265 AVQGLIYVNPPEGNGKPDPTAAARDIRETFORMAMNDEETVALLAGHTTGFCTGAADA 324
QY 285 EKYVGRPAAGAEIEMSLGKNTYGTGHDATITSGLEGAWTKTPTQSNFFENFLGYE 344
Db 325 DN-VGPEPEAGELEGOGLGHNRFSGKAGDTITSGLEVTTWTKTPTQSNDFFEHLGYE 383
QY 345 WELTKSPAGAYQWPKDCAGAGTTPDAHDPSKSHAPFMLTTDLALRMDPDYKISRYYE 404
Db 384 WELTKSPAGAYQWVAKN--AAATIPHAHDPSKLLPMLTSDLALRFPDIYEKISRHFHA 441
QY 405 NPDEFADAFKAWYKLTHRDMGPKVRYLGPVEPOEDLIWQDPIPDVSHPLVDENDIEGLK 464
Db 442 HPDQFADYFARAFKLMHRDMGPRVRYLGPVEVEELIWQDVPKVSFVLVDAQDALLAK 501
QY 465 AKILESGVTSELVSTAWASASTFRNSDKRGANGARIRLAPQKDWENNPQOLARVLKT 524
Db 502 QKISASGLGISQLVSTAWASASTFRGSDKRGANGRLCLAPQSQWEVNPQOLSWLET 561
QY 525 LEGTOEDFNOAQSDNKAVSLADLIVLAGCAGVEKAAKADGHEVOVPPNPGRADATAEOTD 584
Db 562 LRRVQTEFN-AQAGDKRISLADLIVLAGGVGVEQAAKAGAGIVVEVPPFGRTDALQEOTD 620
QY 585 VEAFALEPAADGFRNYIKPEHKVSAEMLVDRQAQLLSAPEMTALVGMNRVLGTNDG 644
Db 621 VSSFALEPFAADGFRNYIKPEHKVSAEMLVDRQAQLLSAPEMTALVGMNRVLGTNDG 680
QY 645 SOHGVTNKPQLSNDFFVNLLDNTKWRASDESKVFEGRDFTKGVKSGTRVDFLIFG 704
Db 681 VKHGVTNKPQLSNDFFVNLLDNTKWRASDESKVFEGRDFTKGVKSGTRVDFLIFG 740
QY 705 SNSELRALAEVYGCADSEKFKVQKAVKAWKMDLDRFDL 744
Db 741 SNSELRALAEVYGCADSEKFKVQKAVKAWKMDLDRFDL 780

RESULT 12
A87626
catalase/peroxidase [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: A87626
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Neilson, K.E.; Elsen, J.; Heidelberg,
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
n, J.; Emolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: A87626
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-737 <STO>
A:Cross-references: GB:AE005673; NID:gl3424687; PIDN:AAK25005.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC3043
C:Superfamily: catalase HPI

Query Match 67.2%; Score 2691; DB 2; Length 737;
Best Local Similarity 69.5%; Pred. No. 2.1e-175;
Matches 508; Conservative 64; Mismatches 145; Indels 14; Gaps 6;
QY 15 NTGKCPFTGGSLSKOSAGGCTKNRQWPNMLNLGILROHSSLSDDNDPFDYAEFKLD 74

```

Db      18 NIAGKCP-----MGHRCGPANRDNWQPSLRLEGLNQHAPSRNPMPGEAFDYABAFKSLD 70
QY      75 LAAYKDLAALMTDSQDWWPADYGHYPFFIRMAHSACTYRIGDGRGGGSGSQRFAPL 134
Db      71 LDVAVSDLHALMTDSQEWNPADPGHVGGLFIRLAWHAAGTYRITDGRGGGSGQRFAPL 130
QY      135 NSHPDNANLDKARLLIWPVKQYGRKISWADLMILGNVALETWNGFKTFGAGGRADWVE 194
Db      131 NSHPDNTLNDKARLLIWPVKQYGALESWADLVVGVNALESNGKTFGAGGRADQWE 190
QY      195 PEDVYWGAEWELGDKRYEGDRELENPLGAVOMGLIYVNPPEGNGKPPDPIAARDIRET 254
Db      191 PEB-LWAGPESTWLDKRRSGRELDPLGAVOMGLIYVNPPEGNGPNPPLASARDIRET 249
QY      255 FGRMANDEETVALIAGGHTFGKTHGAADAETVYVREPAAGIEEMLGKNTYGTGCHA 314
Db      250 FARMANDEETVALIAGGHTFGKTHGAAGDA-SLVGVEPEGALEAGFGWASHGKGTGP 308
QY      315 DTITSGEAWTKYTPQWNSNPFENLFGYEWELTKSPAGAYQWKPKDGAGAGTIPDAHDP 374
Db      309 DAITGGPEVITWQTPTRWSNHFFDNLFKYEWELTQSPAGAKQWAKN--APADIPDAFDP 366
QY      375 SKSHAPFMTTDLALRMDPDYKISRYYENPDEFADAFKAWYKLTHRDMGPKVRYLGP 434
Db      367 NKTHVPRMLTSLALRFDPAYEKISRFFENPDPADAFARAWFKLTHRDMGPIGRYLG 426
QY      435 EYVQEDLIWODPIPVSHPLVDENDIEGLKAKILESLVSELVSTAWASASTFRNSDKR 494
Db      427 LVPKEELIWDPIPAVDHPLADDKIALKAKTLATGLSASDLVSTAWASASTYRQSDKR 486
QY      495 GGANGARIRLAPQKDEVNPNPOOLARVLTLEGIQDFNQADSKDNKAVSLADLIVLAGCA 554
Db      487 GGANGARIRLAPQKDEVNPNPVLAKVLALEGVQDFNMSAGGGKRIADLIVLGAA 546
QY      555 GVEKAADAGHEVQVFPENPGRADATAEQTDVEAFEALPEAADGFRNYIKP-EHKVSAEEM 613
Db      547 AIEKAADAGTSVTVFAPGRMDASAQTDHSAFEALPEPRSDGFRNYRGPGRHYMAPEEA 606
QY      614 LVDRQALLSLAPEMTALVGGMRVLTNYDGSOGHVTNKPQOLSNDFFVNLDLNTRWR 673
Db      607 LVDRQALLSLPELTVLVGLRVLGANADGSKDGVFTNRPGLALSNDFFVNLLSMETTW- 665
QY      674 ASDSKVPEGRDFKTEGVKWCSTRVDLIFGNSSELALAEVYGCADSEKFEKVFKAH 733
Db      666 -SPTANAFAGHRKSSPRWTATRVLDLIFGSHAEALAEVYACADSQEKFEVCDVETAW 724
QY      734 AKYMDLDRFDL 744
Db      725 NKVMNADRLDL 735

RESULT 13
A:0662
N:Alternate names: catalase-peroxidase; hydroperoxidase I; katG catalase-peroxidase bifur
N:Contains: peroxidase (EC 1.11.1.7)
C:Species: Mycobacterium tuberculosis
C:Date: 03-May-1994 #sequence_revision 24-Jul-1998 #text_change 16-Jun-2000
C:Accession: A70519; A40662; S29071; S34036
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
  Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
  Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
  Nature 393, 537-544, 1998
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70519
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-740 <COL>
A:Cross-references: GB:297193; GB:AL123456; NID:g3261816; PIDN:CAB10056.1; PID:g2225952
A:Experimental source: strain H37Rv
R: Heym, B.; Zhang, Y.; Poulet, S.; Young, D.; Cole, S.T.

```

```

J. Bacteriol. 175, 4255-4259, 1993
A:Title: Characterization of the katG gene encoding a catalase-peroxidase required for
A:Reference number: A40662; MUID:93308108; PMID:8320241
A:Accession: A40662
A:Molecule type: DNA
A:Residues: 1-70, 'SRL', 74-170, 'RCARNWASRRSGSAG', 188-192, 'T', 194-207, 'DG', 210-211, 'V',
  'RCRPTS', 610-711, 'APMTR', 711, 'A', 719-721, 'TG', 724-740 <HEY>
A:Cross-references: EMBL:X68081; GB:S42739; NID:g581367
R: Zhang, Y.; Heym, B.; Allen, B.; Young, D.; Cole, S.
  Nature 358, 591-593, 1992
A:Title: The catalase-peroxidase gene and isoniazid resistance of Mycobacterium tuber
A:Reference number: S29071; MUID:92365822; PMID:1501713
A:Accession: S29071
A:Molecule type: DNA
A:Residues: 1-70, 'SRL', 74-80 <zHA>
A:Cross-references: EMBL:X68081
C:Genetics:
A:Gene: katG
A:Start codon: GTG
C:Complex: homotetramer; two non-covalently associated iron protoheme IX groups per t
C:Function: <CAT>
A:Description: as catalase, catalyzes the dismutation of two molecules of hydrogen pe
C:Function: <PEB>
A:Description: as peroxidase, uses hydrogen peroxide to oxidize donor compounds produ
A:Note: active with a broad spectrum of donor compounds, including larger ones which
C:Superfamily: catalase HPI
C:Keywords: chromoprotein; heme; homotetramer; iron; metalloprotein; oxidoreductase
F:108/Active site: His (distal axial ligand) #status predicted
F:270/Binding site: heme iron (His) (proximal axial ligand) #status predicted
F:321/Active site: Trp (tryptophyl radical intermediate) #status predicted

Query Match 63.6%; Score 2544; DB 1; Length 740;
Best Local Similarity 64.3%; Pred. No. 2,2e-165;
Matches 480; Conservative 88; Mismatches 168; Indels 10; Gaps 8;

QY      2 ENHKHSGSYTNTGCKPFTGCGSLKQSGAGTKNRDWWPNMLNGLIRQHSSLSDPND 61
Db      3 EQHPPTTETTTGAASNG-CPVVGVMKYPVEGG--NQDWWPNRLNLKVLHQPAPVADPMG 59
QY      62 POFDYAEFKKDLAAVKDLAALMTDSQDWWPADYGHYPFFIRMAHSACTYRIGDGR 121
Db      60 AAFDYAAEVATIDVDALTRDIEVWTSQVWNPADYGHYPFFIRMAHSACTYRIGDGR 119
QY      122 GGGGSGSQRFAPLNSWPDNANLDKARLLIWPVKQYGRKISWADLMILGNVALETWNGFK 181
Db      120 GGAGGGQRFAPLNSWPDNANLDKARLLIWPVKQYGRKISWADLMILGNVALETWNGFK 179
QY      182 TFCFAGGRADVWPEEDVYWGAEWELGDKRYEGDRELENPLGAVOMGLIYVNPPEGNGK 241
Db      180 TFCFAGGRADVWPEEDVYWGAEWELGDKRYEGDRELENPLGAVOMGLIYVNPPEGNGK 238
QY      242 PDPAAADIRETFFGRMANDEETVALIAGGHTFGKTHGAADAETVYVREPAAGIEEML 301
Db      239 PDPAAADIRETFFGRMANDEETVALIAGGHTFGKTHGAADAETVYVREPAAGIEEML 297
QY      302 LGWKNTYGTGCHGADTTTSGLEGAWTKTPTQWNSNPFENLFGYEWELTKSPAGAYQWKPK 361
Db      298 LGWKNTYGTGCHGADTTTSGLEGAWTKTPTQWNSNPFENLFGYEWELTKSPAGAYQWKPK 357
QY      362 GAGAGTIPDAH-DPSKSHAPFMTTDLALRMDPDYKISRYYENPDEFADAFKAWYKL 420
Db      358 GAGAGTIPDPFGGPGHS--PTMLATDLSLRVDPIYERITRRLWLEHPEELADEFAKAWYKL 415
QY      421 THRDMPGKRYLGPVQPSDLIWDPIVDSVPLVDENDIEGLKAKILESGTIVSELVST 480
Db      416 THRDMPGKRYLGPVQPSDLIWDPIVDSVPLVDENDIEGLKAKILESGTIVSELVST 475
QY      481 AWASASTFRNSDKRGGANGARIRLAPQKDEVNPNPQ-QLARVLTLEGIQDFNQADSKDN 539
Db      476 AWASASTFRNSDKRGGANGARIRLAPQKDEVNPNPQ-QLARVLTLEGIQDFNQADSKDN 535
QY      540 KAVSLADLIVLAGCAVEKAAKADAGHEVQVFPENPGRADATAEQTDVEAFEALPEAADG 599

```

[illegible]

309 SSYGTSGKDAITSGLEVYVMTPTTKWDSNFFLETLGYEWELTKSPAGAWQFTAKDGAGA 368

366 GTIPDAHDPSKSHAPMLTTDLALRMDPDVEKISRYYENPDDEFADAFKAWKLTTHRDM 425

369 GTIPDPFG-CAGRAPMLVTDISLRSPYADITRWLDHPPELADAFKAWKLTTHRDM 427

426 GPKVRYLGPVPEQDILWQDPIPDVSHPLVDENDIEGLKAKILKESGLTVSELYSTAWASA 485

428 GPTSRYLGPVWABPQL-WQDPVPAVDHVLVDNDVAALKKKVLDSGLSPQLVKVTAWSAA 486

486 STFRNSDKKGGANGARILAKQDKWENVNPPQOLARVYLKLEGTOEDFNQAQSDNKAYSLA 545

487 ASYRNTDKKGGANGKRLRQPKRSWEVNSPELKVLPVLEKIQDDFNASASGGKKRISA 546

546 DLIVLAGCAGVEKAAKDAGHEVOVPPNPGRADATBQTDVEAFEALEPAADGFRNYIKPE 605

547 DLIVLAGSAVEKAAKDAGYEISVHPFAPGRTDASQESTDVSFAVLEPRADGFRNYIRG 606

606 HKVSAEMLVDRAQLLSAPENTALVGMNRVLGTYNGDSQHGVTFTKPGQLSNDFVNL 665

607 EKAPLEQLLTERAYLLGVTPGENTVLVGGILRAGLNGHSSKHGVFTRDPCALTNDFVNL 666

666 LDLNTKWRASDESDDKVFEGRDFKTEGVKWSGRFYDLIFGNSNELRALAEVYGCADSEERF 725

667 LDMGTWKASETAENYEGGRASGALKWTATANDLVFGNSVYLGVEVYTAQDDAHGKF 726

726 VKDFVKAWAKVMDLDRFDLK 745

727 VEDFVAWVKVMSDRFDLK 746

RESULT 15

AI0936

catalase (hydroperoxidase 1) [imported] - Salmonella enterica subsp. enterica serovar

C:Species: Salmonella enterica subsp. enterica serovar typhi

A:Note: this species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: AI0936

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr

, S.; Moulle, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AI0936

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-726 <FAR>

A:Cross-references: GB:AL513382; PIDN:CAD09515.1; PID:g16504632; GSPDB:GN00176

C:Genetics:

A:Gene: SVY3760

C:Superfamily: catalase HPI

Query Match 61.1%; Score 2444.5; DB 2; Length 726;

Best Local Similarity 63.6%; Pred. No. 1.3e-158;

Matches 468; Conservative 83; Mismatches 166; Indels 19; Gaps 7;

QY 11 TYNTNTGGKCPFTGGSLKSGAGGTTKNRWPNMLNLGIIRQHSSISDPNDPFDYAEFP 70

Db 7 THNTLTGKCPFFGGHDSRAGTAGTASRWMPNQLRVDDLNNQHSNPNLGEDYRKEF 66

QY 71 KKDLAAVKKDLAALMTDSDQWMPADYGHVGPFFIRMAWHSAGTYRIGRGGGSGSOR 130

Db 67 SKLDYSAKGLDKALLTDSQWMPADQWGSVGLFIRMAWHGAGTYRISDGRGAGGQQR 126

QY 131 FAPLNSWPONANLDKARLLWP1KQYGRKISWADLMILTGNVALETMGFKYFGFAGGGA 190

Db 127 FAPLNSWPONVSLDKARRLLWP1KQYGGKISWADLFILAGNVALENSGFRYFGCAGRE 186

QY 191 DWMEPEDVYWGAEWMLGDKRYEGDELENPLGAVOMGLIYVNPBGPKPDPIAARD 250

Db 187 DWMEPDLDVNWGDEKAWLTHRHPEA--LAKAPLGATEMGLIYVNPBGHSGEPLSAAA 244

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 7, 2003, 19:38:38 ; Search time 74.5995 Seconds
(without alignments)
2577.085 Million cell updates/sec

Title: US-09-884-889-8
Perfect score: 4002

Sequence: 1 MHNKSGSGSYNTGK.....VKDFVKAQKVMLDREDFLK 745

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	2989.5	74.7	736	17 Q8TS34	Q8ts34 methanosarc
2	2884.5	72.1	756	16 Q987S0	Q987s0 rhizobium l
3	2867	71.6	748	2 Q939D2	Q939d2 burkholderi
4	2787.5	69.7	727	16 Q92XG8	Q92xg8 rhizobium m
5	2782.5	69.0	756	16 Q8PMX4	Q8pmx4 xanthomonas
6	2759.5	69.0	735	16 Q9KDE6	Q9kee6 bacillus ha
7	2714.5	67.8	743	2 Q9S5K5	Q9s5k5 bacillus st
8	2714.5	67.8	743	2 Q9S5X0	Q9s5x0 bacillus st
9	2714.5	67.8	744	2 Q9S5N2	Q9s5n2 bacillus st
10	2714.5	67.8	744	2 Q9S5K7	Q9s5k7 bacillus st
11	2714.5	67.8	745	2 Q9S5Q5	Q9s5q5 bacillus st
12	2714.5	67.8	746	2 Q9S5N4	Q9s5n4 bacillus st
13	2714.5	67.8	748	2 Q9S5L8	Q9s5l8 bacillus st
14	2714.5	67.8	749	2 Q9S5P6	Q9s5p6 bacillus st
15	2714.5	67.8	750	2 Q9S5Q6	Q9s5q6 bacillus st
16	2714.5	67.8	750	2 Q9S5Q7	Q9s5q7 bacillus st

17	2714.5	67.8	751	2 Q9S5P7	Q9s5p7 bacillus st
18	2714.5	67.8	751	2 Q9S5L5	Q9s5l5 bacillus st
19	2714.5	67.8	751	2 Q9S5N9	Q9s5n9 bacillus st
20	2714.5	67.8	751	2 Q9S5P1	Q9s5p1 bacillus st
21	2714.5	67.8	751	2 Q9S5L1	Q9s5l1 bacillus st
22	2714.5	67.8	751	2 Q9S5M4	Q9s5m4 bacillus st
23	2714.5	67.8	751	2 Q9S5M1	Q9s5m1 bacillus st
24	2714.5	67.8	751	2 Q9S5Q8	Q9s5q8 bacillus st
25	2714.5	67.8	751	2 Q9S5P9	Q9s5p9 bacillus st
26	2714.5	67.8	751	2 Q9S5N0	Q9s5n0 bacillus st
27	2714.5	67.8	751	2 Q9S5N3	Q9s5n3 bacillus st
28	2714.5	67.8	751	2 Q9S5M0	Q9s5m0 bacillus st
29	2714.5	67.8	751	2 Q9S5K4	Q9s5k4 bacillus st
30	2714.5	67.8	751	2 Q9S5P2	Q9s5p2 bacillus st
31	2714.5	67.8	751	2 Q9S5Q0	Q9s5q0 bacillus st
32	2714.5	67.8	751	2 Q9S5N6	Q9s5n6 bacillus st
33	2714.5	67.8	751	2 Q9S5L7	Q9s5l7 bacillus st
34	2714.5	67.8	751	2 Q9S5Q3	Q9s5q3 bacillus st
35	2714.5	67.8	751	2 Q9S5M7	Q9s5m7 bacillus st
36	2714.5	67.8	751	2 Q9S5M8	Q9s5m8 bacillus st
37	2714.5	67.8	751	2 Q9S5L6	Q9s5l6 bacillus st
38	2714.5	67.8	751	2 Q9S5L9	Q9s5l9 bacillus st
39	2714.5	67.8	751	2 Q9S5L3	Q9s5l3 bacillus st
40	2714.5	67.8	751	2 Q9S5Q1	Q9s5q1 bacillus st
41	2714.5	67.8	751	2 Q9S5Q4	Q9s5q4 bacillus st
42	2714.5	67.8	751	2 Q9S5L2	Q9s5l2 bacillus st
43	2714.5	67.8	751	2 Q9S5P8	Q9s5p8 bacillus st
44	2714.5	67.8	751	2 Q9S5M2	Q9s5m2 bacillus st
45	2714.5	67.8	751	2 Q9S5N8	Q9s5n8 bacillus st

ALIGNMENTS

RESULT 1

Q8TS34
ID Q8TS34 PRELIMINARY; PRT; 736 AA.
AC Q8TS34;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Catalase/peroxidase.
GN MA0972.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Allen N., Navlor J., Stange-Thomann N., DeArellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Graham D.A., Guss A.M.,
RA Hederich R., Ingram-Smith C., Kuettnar H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Unayam L.A., White O., White R.H., de Macario E.C.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
and physiological diversity.";
RL Genome Res. 12:532-542(2002).
DR EMBL; AE010762; AAM04404.1; -;
DR InterPro; IPR000763; Bac_ctadase/prase.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00141; peroxidase; 1.
DR TIGRfams; TIGR00198; cat_per_HPI; 1.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS00436; PEROXIDASE_2; 1.
KW Peroxidase; Complete proteome.

SQ SEQUENCE 736 AA; 81051 MW; 67A0508CD91BFA7A CRC64;

Query Match 74.7%; Score 2989.5; DB 17; Length 736;
Best Local Similarity 76.3%; Pred. No. 9e-185;
Matches 551; Conservative 62; Mismatches 108; Indels 1; Gaps 1;

QY 23 TGSLSKQSGAGGKTRDWPWNLNLTGILQHSLSLSDPDDFYAEFFKDLAAVKKDL 82
DB 15 TSGANKQETGRDMSNRDWPWNLNLTGILQHSLSLSDPDDFYAEFFKDLAAVKKDL 74

QY 83 AALMTDSQDWPADYGHYGPETIRMAHWSAGTYRIGDGRGGGSGSOREAPLNSWPDNAN 142
DB 75 AALMTDSQDWPADYGHYGPETIRMAHWSAGTYRIGDGRGGGSGSOREAPLNSWPDNAN 134

QY 143 LDKARLLLPKQYGRKISWADLMILTNVALETMGFTTFGAGRADYWEEDVYWG 202
DB 135 LDKARLLLPKQYGRKISWADLMILTNVALETMGFTTFGAGRADYWEEDVYWG 194

QY 203 AETEWLGDYRIGDRELEPLGAVQMLIYVNPPEGNGKPDPTIAAADRIETFGRMAMD 262
DB 195 SEDTWLGDYRIGDRELEPLGAVQMLIYVNPPEGNGKPDPTIAAADRIETFGRMAMD 254

QY 263 EETVALIAGHTFGKTHGAADAEKYYGEPAAAGIEEMSLGWKNTYGTGAGDATTISGLE 322
DB 255 EETVALIAGHTFGKTHGAADAEKYYGEPAAAGIEEMSLGWKNTYGTGAGDATTISGLE 313

QY 323 GAWTKPTQWNSNFFNLFGYEWELTKSPAGAYQWKPDGAGAGTIPDAHDPSKSHAPFM 382
DB 314 VTWNTPTQWNSNFFNLFGYEWELTKSPAGAYQWKPDGAGAGTIPDAHDPSKSHAPFM 373

QY 383 LTTDLALRMDPDYKISRBYENPDEFADAFKAWKLTTRDMGPKVRYLGPVPODILI 442
DB 374 MTDLALRMDPDYKISRBYENPDEFADAFKAWKLTTRDMGPKVRYLGPVPODILI 433

QY 443 WDDPDPVSHPLVDENDIEGLAKILSEGLTSELVSTAWASSTERNSDKRGANGARI 502
DB 434 WDDPDPVSHPLVDENDIEGLAKILSEGLTSELVSTAWASSTERNSDKRGANGARI 493

QY 503 RLAPQKDWNNPQOLARVLKTLGTEQDFNQASDNKAVSLADLIVLAGCAGVEKAAKD 562
DB 494 RLAPQKDWNNPQOLARVLKTLGTEQDFNQASDNKAVSLADLIVLAGCAGVEKAAKD 553

QY 563 AGHEVOVPNGRADATAEQTVAEPAEPAADGPNYKPKHKYSAEMLYDRAQLIS 622
DB 554 AGHEVOVPNGRADATAEQTVAEPAEPAADGPNYKPKHKYSAEMLYDRAQLIS 613

QY 623 LSAPETALVGGMRVLTGDSHGVTNKKPQLSNDFFVNLIDLNTKWRASDESDFK 682
DB 614 LSAPETALVGGMRVLTGDSHGVTNKKPQLSNDFFVNLIDLNTKWRASDESDFK 673

QY 683 EGRDFTGKVGSGTRVDLIFGNSSELRAEYVGCADSEKFKVDYKAWKVMOLDRF 742
DB 674 EGRDFTGKVGSGTRVDLIFGNSSELRAEYVGCADSEKFKVDYKAWKVMOLDRF 733

QY 743 DL 744
DB 734 DL 735

RESULT 2
Q987S0 PRELIMINARY; PRT; 756 AA.

AC Q987S0;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Catalase/peroxidase.
GN MLR6940.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]

RP SEQUENCE FROM N.A.
RX STRAIN=MAFF303099;
RC MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT *Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.*;
RL DNA Res. 7:331-338(2000).
DR ENBL; AF003010; BAB53130.1; -.
DR HSSP; P00431; 1BES.
DR InterPro; IPR000763; Bac_ctase/prase.
DR InterPro; IPR002016; peroxidase.
DR Pfam; PF00141; peroxidase; 1.
DR PRINTS; PR00458; PEROXIDASE.
DR TIGRFAMs; TIGR00198; cat_per_HPI; 1.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS00436; PEROXIDASE_2; 1.
KW Peroxidase; Complete proteome.
SQ SEQUENCE 756 AA; 82348 MW; 7B8BAFC5DC5900AC CRC64;

Query Match 72.1%; Score 2884.5; DB 16; Length 756;
Best Local Similarity 71.7%; Pred. No. 5.8e-178;
Matches 533; Conservative 74; Mismatches 129; Indels 7; Gaps 2;

QY 2 ENKHSGSSTYNTGKCPFTGGSLKQSGAGGKTRDWPWNLNLTGILQHSLSLSDPND 61
DB 19 QNRSDDMDAKTDNSAGKCPVAHGS-----AGRTNRDWPWNLNLTGILQHSLSLSDPND 72

QY 62 PDFDYAEFFKDLAAVKKDLAALMTDSQDWPADYGHYGPETIRMAHWSAGTYRIGDGR 131
DB 73 EAPDYAEFFKDLAAVKKDLAALMTDSQDWPADYGHYGPETIRMAHWSAGTYRIGDGR 132

QY 122 GGGSGSOFAPLNSWPDNANLTKARLLLPKQYGRKISWADLMILTNVALETMGPK 101
DB 133 GGAGAGQORFAPLNSWPDNANLTKARLLLPKQYGRKISWADLMILTNVALETMGPK 192

QY 182 TFGAGGADYWEEDVYWGAEETWLGDKRYEGDRELEPLGAVQMLIYVNPPEGNGK 241
DB 193 TFGAGGADYWEEDVYWGAEETWLGDKRYEGDRELEPLGAVQMLIYVNPPEGNGK 252

QY 242 PDPIAARDIETFGRMAMDDEETVALIAGHTFGKTHGAADAEKYYGEPAAAGIEEMS 301
DB 253 PDPIAARDIETFGRMAMDDEETVALIAGHTFGKTHGAADAEKYYGEPAAAGIEEMS 311

QY 302 LGWKNTYGTGCHGADTTISGLEGAWTKPTQWNSNFFNLFGYEWELTKSPAGAYQWKPK 361
DB 312 LGWKNTYGTGCHGADTTISGLEGAWTKPTQWNSNFFNLFGYEWELTKSPAGAYQWKPK 371

QY 362 GAGAGTIPDAHDPSKSHAPFMLTDLALRMDPDYKISRBYENPDEFADAFKAWKLT 421
DB 372 GAGAGTIPDAHDPSKSHAPFMLTDLALRMDPDYKISRBYENPDEFADAFKAWKLT 431

QY 422 HRDMGPKVRYLGPVPODILIMQDPTPDVSHPLVDENDIEGLAKILSEGLTSELVST 481
DB 432 HRDMGPKVRYLGPVPODILIMQDPTPDVSHPLVDENDIEGLAKILSEGLTSELVST 491

QY 482 WASASTFRNSDKRGGANGARILAPQKDWNNPQOLARVLKTLGTEQDFNQASDNKA 541
DB 492 WASASTFRNSDKRGGANGARILAPQKDWNNPQOLARVLKTLGTEQDFNQASDNKA 551

QY 542 VSLADLIVLAGCAGVEKAAKADAGHEVQVFPNPGRADATAEQTVAEPAEPAADGPNY 601
DB 552 VSLADLIVLAGCAGVEKAAKADAGHEVQVFPNPGRADATAEQTVAEPAEPAADGPNY 611

QY 602 IRPEHKVSAEMLYDRAQLISAPETALVGGMRVLTGDSHGVTNKKPQLSNDFFVNLIDL 661
DB 612 IRPEHKVSAEMLYDRAQLISAPETALVGGMRVLTGDSHGVTNKKPQLSNDFFVNLIDL 671

QY 662 FVNLIDLNTKWRASDESDFKVDLIFGNSSELRAEYVGCADSEKFKVDYKAWKVMOLD 721
DB 671 FVNLIDLNTKWRASDESDFKVDLIFGNSSELRAEYVGCADSEKFKVDYKAWKVMOLD 731

```
Db 672 FVNLLDMGTENKATSDAKDVFEGDRDKTGCVKWTGTRADLIFGSHSQLRALAEVYATADA 731
QY 722 EEKFKVQFVKAWKVMQDLDRDL 744
Db 732 KAKFAKDFVAVWTKVMNADREDI 754

RESULT 3
Q939D2 PRELIMINARY; PRT; 748 AA.
AC Q939D2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Catalase-peroxidase protein katG.
GN katG.
OS Burkholderia pseudomallei (Pseudomonas pseudomallei).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=28450;
RN [1]
RP SEQUENCE FROM N.A.
RA Loprasert S., Sallabhan R., Whangsk W., Mongkolsuk S.;
RT "A Burkholderia pseudomallei oxyR mutant has altered biofilm
RT formation, autoaggregation, extracellular protease activity, and
RT oxidative stress response.";
RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AY040244; AAK7426.3; -.
DR HSSP; P00431; 1BES.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00141; peroxidase; 1.
DR PRINTS; PR00458; PEROXIDASE.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS00436; PEROXIDASE_2; 1.
KW Peroxidase.
SQ SEQUENCE 748 AA; 81760 MW; F59D6C85A6850AF3 CRC64;

Query Match 71.6%; Score 2867; DB 2; Length 748;
Best Local Similarity 74.0%; Pred. No. 7.7e-177;
Matches 541; Conservative 64; Mismatches 112; Indels 14; Gaps 6;

QY 19 KCPFTGSLKQASGGTKNRDWPNNMLNLGILRHSSLSNDPDPDYAEFFKKLDLAAY 78
Db 26 KCPE-----HQAAGNTSNRDWPNQLDLSILHRHSSLSDPGKDFNTAQAFKLDLAAY 80
QY 79 KDLAALMTDSQDWNPADYGHYGFPIFIRMAHWSAGTYRIGDGRGGSGSQRFAPLNSWP 138
Db 81 KRDLEALMTTSDQWNPADFGHYGLFIRMAHWSAGTYRTADGRCGAGGEGQQRFAPLNSWP 140
QY 139 DNANLDKARLLLPITKQYGRKISWADLMILTGNVALETMTGKTFGAGGRADVWEPEED 198
Db 141 DNANLDKARLLLPITKQYGRAISWADLIILITGNVALESMTGKTFGAGGRADVWEPEED 199
QY 199 VYWGAEETWL-----GDKRYGDELEPLNGLAVQMLIYNVPEGNPGKDPDTAAARDIRE 253
Db 200 VYWGSEKILWELSGSPNSRYSGDQLENPLAQLVQMLIYNVPEGPDGPDVPAARDIRD 259
QY 254 TFGRMAMDEETVALIAGCTFGKTHGAADAPEKYVYGREPAAGTEMSLGNKTYGTGTHG 313
Db 260 TFGRMAMDEETVALIAGCTFGKTHGAGPASN-VGAEPGAAGTEAQLGWSAYRTGKG 318
QY 314 ADTITSGLEGATWTKPTQWSNNFENLFGYEWELTKSPAGAYQWKPKDGAGAGTIPDAHD 373
Db 319 ADATITSGLEVTTTPTQWSNNFENLFGYEWELTKSPAGAHQWYAK-GADA-VTPDAFD 376
QY 374 PSKHAPMLTDLALMDPPEYKISRYRYENPEDEFAFAKAKLTHRDMPKRVRYLG 433
Db 377 PSKHURPMLTDLSLRDPAYEKISRPFENPEQFADAFKAMFKLTHRDMPKRVRYLG 436
QY 434 PEVQEDLIWQDPIPDVSHPLVDENIEGLKAKILESGLTVSELVSTAWASASTPRNSDK 493
Db 437 PEVPAEVLWQDPIPAVDHPLIDAADAAELKAKVLASGLTVSLSQVSTAWAASSTFRGSDK 496
```

```
QY 494 RGGANGARIRLAPQKDEVNPNPQQLARVLKLTLEGTOEDFNOAQSDNKAVSLADLIVLAGC 553
Db 497 RGGANGARIRLAPQKDEWANOPEQLAAVLETLEAIRTAFNGAQGGCKQVSLADLIVLAGC 556
QY 554 AGVEKAADAGHEVOVPENPGPADATAGOTDVEAFEALPEADGFRNFKPEHKVSAEEM 613
Db 557 AGVEQARAGHNAVTPFAPGRDASQEQOTDVESMAVLEPVADGFRNFKGIRYPAEVL 616
QY 614 LVDRQAQLLSLAPENTALVGGMRVLTGNYDGSQHGVTNKPQOLSNDFVNLILDLNTKWR 673
Db 617 LVDRQAQLLSLAPENTALVGGMRVLTGNYDGSQHGVTNKPQOLSNDFVNLILDLNTKWR 676
QY 674 ASDESDKVFEGDRDFTGEVKSQGTVDLIFGNSERLALAEVYGCADSEKFKVDFKAW 733
Db 677 PTAADAVFEGDRDRTGELKWTGTRVDLVFGSHSQLRALAEVYGSADAQGEKVRDFVAVW 736
QY 734 AKVMNLDLDRFDL 744
Db 737 NKVMNLDLDRFDL 747

RESULT 4
Q92XG8 PRELIMINARY; PRT; 727 AA.
AC Q92XG8;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Catalase/peroxidase (EC 1.11.1.6).
GN RAI286 OR SMA2379.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid psyma (megaplasmid 1).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Garjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSyma megaplasmid.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
DR EMBL; AE007314; AAK6594.1; -.
DR HSSP; P00431; 1BES.
DR InterPro; IPR00763; Bac_ctase/prase.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00141; peroxidase; 1.
DR PRINTS; PR00458; PEROXIDASE.
DR TIGRFAMs; TIGR00198; cat_per_HPI; 1.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS00436; PEROXIDASE_2; 1.
KW Peroxidase; Oxidoreductase; Plasmid; Complete proteome.
SQ SEQUENCE 727 AA; 80292 MW; 31086D8C7AA793DF CRC64;

Query Match 69.7%; Score 2787.5; DB 16; Length 727;
Best Local Similarity 71.1%; Pred. No. 1e-171;
Matches 521; Conservative 72; Mismatches 127; Indels 13; Gaps 6;

QY 14 TMTGKCPFTGSLKQASGGTKNRDWPNNMLNLGILRHSSLSNDPDPDYAEFFKKL 73
Db 5 SDSAGKCP-----VAHTAPRGSRNRDWPDLQDVVLRHSHSLDPLGNTFNFAEKKL 59
QY 74 DLAAVKDLAALMTDSQDWNPADYGHYGFPIFIRMAHWSAGTYRIGDGRGGSGSQRFAP 133
Db 60 DLDALKRDLRALMTDSQDWNPADFGHYGLFIRMAHWSAGTYRTDGRGAGGQGRFAP 119
QY 134 LNSWPDNANLDKARLLLPITKQYGRKISWADLMILTGNVALETMTGKTFGAGGRADVW 193
```

Db 120 LNSWPNANLDKARRLLWPIKQKYGNRISWADLLILGNVALESKMGKTFGAGGRVDVW 179

Qy 194 EPEEDVYWGAEWELGDKRYEGDRELENPLGAVOMGLIYVNEPGNGKPPDIAAARDIRE 253

Db 180 EPEE-LFWGPEGTWLGDSYGERQISLEPLAAVOMGLIYVNEPGNGPPVAAARDIRE 238

Qy 254 TFGRMANDEETVALIAGHTFGKTHGAAADAERYVGPAAAGTEEMSLGKWNKYTGTHG 313

Db 239 TFRMANDEETVALIAGHTFGKTHGAGD-PSFIGADPEGGATEDDGLGKSTFGTGVG 297

Qy 314 ADRTISLEGAWKTPQWNNFENLFGYEWELTKSPAGAYQWKPDGAGAGTIPDAHD 373

Db 298 KDAITGSGPEVTSOTPTWNSHFFENLFWHEWELTKSPAGAYQWKA--AEATIPDAYD 355

Qy 374 PSKSHAPFMTLTDIALRMDPDYKISRYYENDEFAFAKAWYKLTNRDMGPKVRYLG 433

Db 356 PSRKHVPTRLTDLSDLPAYEKISRFLNDEFAFAKAWYKLTNRDMGPKVRYLG 415

Qy 434 PEVPOEDLIWQDPTPDVSHPLVDENDIEGLKAKITLESGLTVSELVSTAWASASTFRSDK 493

Db 416 PEVPAEDLIWQDPTPDVSHPLVDENDIEGLKAKITLESGLTVSELVSTAWASASTFRSDK 475

Qy 494 RGGANGARIRLAPQKOWEVRNPAQLARVLSVLEGIQDFNAQTDGKKISLADLIVLAGG 553

Db 476 RGGANGARIRLAPQKOWEVRNPAQLARVLSVLEGIQDFNAQTDGKKISLADLIVLAGG 535

Qy 554 AGYEKAADAGHEVQVPPNFGRADATAEQTDVFAFEALPEAAGDFRNYIKPEHK--VSAE 611

Db 536 AAYEKAADAGGHDITVFTPCRDMSAOTDAASFALEPRADGFRNYVSTTRQCFMKPE 595

Qy 612 EMLVDRAQLLSLAPEMTALVGMGRVLTNYDGSOGHGVETNKPQSLNSDFVNLDLNTK 671

Db 596 EALVDRAQLLTFAPETMTLVGLRLVLAG--EPKHGVFTSRPEALTNDFVNLDMGTQ 653

Qy 672 WRASDESDEKVFEGRDKFTGKWSGTRVDLIFGSNSLRALAEVYGADSEKFEVKDFVK 731

Db 654 WSPIEGEGYEGRDRRTGAARTGTIRVDLIFGSHSOLRAFAEYVAGSDAREKVFVDFA 713

Qy 732 AWAKVMDLDRFDL 744

Db 714 AMTKVMNADRFDL 726

RESULT 5

Q8PMX4 ID Q8PMX4 PRELIMINARY; PRT; 756 AA.

AC Q8PMX4; AC

DT 01-OCT-2002 (T-EMBLrel. 22, Created)

DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)

DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)

DE Catalase.

GN KATG OR XAC1301.

OS Xanthomonas axonopodis (pv. citri).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;

OC Xanthomonadaceae; Xanthomonas.

OX NCBI_TaxID=92829;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-306 / ATCC 13902 / XV 101;

RX MEDLINE-2022145; PubMed-12024217;

RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,

RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,

RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,

RA Camarotte G., Canhavan F., Cardoso J., Chamergo L.P.,

RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,

RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,

RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,

RA Katsuyama A.M., Kish L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,

RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,

RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,

RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,

RA Perreira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,

RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,

RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,

RA Setubal J.C., Kitajima J.P.;

RT "Comparison of the genomes of two Xanthomonas pathogens with differing

RT host specificities";

RL Nature 417:459-463(2002).

DR EMBL; AE011760; AAM36172.1; -.

DR InterPro; IPR000763; Bac_cbase/prase.

DR InterPro; IPR002016; Peroxidase.

DR Pfam; PF00141; peroxidase; 1.

DR PRINTS; PR00458; PEROXIDASE.

DR TIGRFAMs; TIGR00198; cat_per_HPI; 1.

DR PROSITE; PS00435; PEROXIDASE_1; 1.

DR PROSITE; PS00436; PEROXIDASE_2; 1.

KW Complete proteome.

SQ SEQUENCE 756 AA; 82827 MW; E7B2FDD5E979CF0F CRC64;

Query Match 69.08; Score 2762.5; DB 16; Length 756;

Best Local Similarity 67.38; Pred. No. 4.5e-170;

Matches 513; Conservative 85; Mismatches 131; Indels 33; Gaps 6;

Qy 6 HSGSTTYNTNTGKCPFTGSGSKQAGGKTKNRDWPMLNMLGILRQHSLSLDPNDPFD 65

Db 4 HTGIA---MTTEAKCPFNHAVV---GTGTTNRDWPMLNMLGILRQHSLSLDPNDPFD 56

Qy 66 YAEFEKLDLAAYVKDLAALMTDSQDWNPADYGHYGFPTFMWHSAGTYRIGDGRGGG 125

Db 57 YADAFKHLDLQALKDLHALMTDSQDWNPADYGHYGFPTFMWHSAGTYRIGDGRGGG 116

Qy 126 SGOFRFAPLNSWPDNANLKLWPKIKYKISKWADLMILTNVALETGKTFGPG 185

Db 117 RGOFRFAPLNSWPDNANLKLWPKIKYKISKWADLMILTNVALETGKTFGPG 176

Qy 186 AGGRADYWEPEEDVYWGAEWEL--GDKRYE-----GPRELENP 222

Db 177 AGGREDTWEPDQDLTWGRETAKWLGDERYSRSGVDEAHVLYKDDSQVPHTRDLENP 236

Qy 223 LGAVOMGLIYVNEPGNGKPPDIAAARDIRETFGRMANDEETVALIAGHTFGKTHGAA 282

Db 237 LAAVOMGLIYVNEPGNGKPPDIAAARDIRETFGRMANDEETVALIAGHTFGKTHGAG 296

Qy 283 DAERYVGREPAAAGIEEMSLGWNKYCTGTHGADTITSGEGAWTKTPTONSNFFENLFG 342

Db 297 PAD-YVGAEPAGELESQSGFGWHNRYSGKGAUTITSGLEVTWTTTTPAQNSDYFDHLFG 355

Qy 343 YEWELTKSPAGAYQWKPDGAGAGTIPDAHDPSKSHAPFMTLTDIALRMDPDYKISRYY 402

Db 356 FEWELSKSPAGAHQWAKN--ADAIIPDAHDASKRHRPTMLTDLALUREDPAYEASIRRF 413

Qy 403 YENPDEPADAFAKAWYKLTNRDMGPKVRYLGPEVPOEDLIWQDPTPDVSHPLVDENDIEG 462

Db 414 QOHPEQPADAFARAWFKLTNRDMGPRARYLGADVPAAELVMDQDPVPAVDHALVDAQDAAA 473

Qy 463 LKAKILFSLTVSELVSTAWASASTFRNSDKRGANGARIRLAPQKOWEVRNPAQLARVLS 522

Db 474 LKOTILASGLSVAILVSTAWASASTFRNSDKRGANGARIRLAPQKOWEVRNPAQLARVLS 533

Qy 523 KTLLEGIOEDFNQAQSDNKAIVSLADLIYLAGCAGYKAAKADAGHEVQVPPNPGRADATAEQ 582

Db 534 ATLERIQADFNAAQSGGKKISLADLIYLAGCAVEAAQAAAGHQVTPVPFAPGRDASQEQ 593

Qy 583 TDYEAFALPEADGDFRNYIKPEHKVSAEMLVDRAQLLSLAPEMTALVGMGRVLTNY 642

Db 594 TDYESFAVLEPVDGFRNFRARYVPAEALLDKAOLLTLTAPELTVLGGRLVGANV 653

Qy 643 DGSOGHGVETNKPQSLNSDFVNLDLNTKWRASDESDEKFEVKDFGKTFGKWSGTRVDLI 702

Db 654 GDSKHGVFTSRPGVLSNDFFNALDMRTKATSEAKVEFEGHRSRSTGELRWTGTRVDLY 713

Qy 703 FGSNSILRALAEVYGCADSEKFEVKDFGKTFGKWSGTRVDLI 744

Db 714 FGSNSILRALAEVYGCADSEKFEVKDFGKTFGKWSGTRVDLI 755

Db 480 D K R G A N G A R I R L A P O R D W E V N Q P E Q L E K V L S V L E N I Q S Q L ----- D K K V S I A D L I V L G 533

Db 232 KPDPKAAARDIRETFRMGNDDEETVALIAGGHTFGKAHGAP-THVGPEPEAPTEAQ 290
Qy 301 SLGKNTYGTGHGADTTTSGLEGAWTKPTQWNSNFFENLFGYEWELTKSPAGAYONKPK 360
Db 291 GLGWISSYKKGKSDTTTSGLEGAWTKPTQWNTSYFDMFLGYDWMWLTSPAGAWQNAV 350
Qy 361 DGAGAGTIPDAHPSKSHAPFMTLTDLALRMDPDYEKISRYYENPDEPDAFAKAWYKL 420
Db 351 DPDEKDLAPDAEDPSKKVPTMMWTTDLALRFDPEYKIRRFHQNPPEFAFAFAWFKL 410
Qy 421 THRDMPKRYLGPVEPOEDLIWQDPIPDVSHPLVDENDIEGLKAKLTESGLTVSELVST 480
Db 411 THRDMPKRYLGPVEPKDFIWDPIPEVDYELT-EAEIEIKAKLTNSGLTVSELVST 469
Qy 481 AWASASTFRNSDRKGGANGARIRLAPOKDWEVNNPOOLARVLTLEGIQEDFNOAQSDNK 540
Db 470 AWASASTFRNSDRKGGANGARIRLAPOKDWEVNEPERLAKVLSVYEDIQREL-----PK 523
Qy 541 AVSLADLIVLAGCAGVEKAAKADAGHEVOVPFNCRADATAEQTQDVEAFEALEPAADGFRN 600
Db 524 KVSADIIVLGGSAAVEKAARDAGFVKVFPFGRGDQEQTDVSEFAVLEPFDGFRN 583
Qy 601 YIRPEKVSABEMLVDRAQLLSAPENTALVGMVRLGTYNGDSHGCVFTNKPGLSND 660
Db 584 YRQEYSVPPPELLVDKQALLGLTAPENTVLGGLRVLGANYRDLPHGVFTDRIGVLTND 643
Qy 661 FFVNLLDLNTKWRASDESDFEGRDPKTGEVKWSGTRVDLIFGNSSELRALAEVYGCAD 720
Db 644 FFVNLLDMNYEWVPTDSG--IYEIRDRKTGEVWTRATRVLDLIFGNSILRSYAEFYAQQD 701
Qy 721 SEEFVKDFVKAQKAVMDLDRFDL 744
Db 702 NQEFVRDFINAWVKVNNADRFDL 725

RESULT 8

Q955R0 ID Q955R0 PRELIMINARY; PRF; 743 AA.
AC Q955R0; DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Peroxidase/catalase (EC 1.11.1.6) (Catalase-peroxidase).
GN CAT.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RA Matsuura T., Miyai K., Trakulnaleaamsai S., Yomo T., Shima Y.,
RA Miki S., Yamamoto K., Urabe I.;
RT "Evolutionary molecular engineering by random elongation
mutagenesis";
RL Nat. Biotechnol. 17:58-61(1998).
CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) -> O(2) + 2 H(2)O.
CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL
CC PEROXIDASE/CATALASE SUBFAMILY.
DR EMBL; AB020120; BAA37032.1; -;
DR HSSP; P00431; 1BJ9.
DR InterPro; IPR000763; Bac_ctase/prase.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00141; peroxidase; 1.
DR PRINTS; P00458; PEROXIDASE.
DR TIGRFAMs; TIGR00198; cat_per_HPT; 1.
DR PROSITE; PS00435; PEROXIDASE.1; 1.
DR PROSITE; PS00436; PEROXIDASE.2; 1.
KW Heme; Hydrogen peroxide; Iron; Oxidoreductase; Peroxidase.
SQ SEQUENCE 743 AA; 84078 MW; A2C54225AALCSB6F CRC64;
Query Match 67.8%; Score 2714.5; DB 2; Length 743;
Best Local Similarity 68.7%; Pred. NO. 5.5e-167;
Matches 511; Conservative 75; Mismatches 139; Indels 19; Gaps 6;

Qy 1 MENKHSSTYNTNTGKCPFTGGSLSKQAGGKTNRDMPNMLNTGILRQHSLSLDPN 60
Db 1 MENQ-----NRQNAACQPFHGSVTNOSS-NRTTNKDWPNOLNLSILHDKRKNPH 51
Qy 61 DPFDYAEPEKDLAALMTDSQDWWPADYGHYOPFFIRMAHWSAGTYRGDG 120
Db 52 DEEFYAEPEKDLKLTMTESQDWWPADYGHYOPFFIRMAHWSAGTYRGDG 111
Qy 121 RGGGSGSQFAPNSNPDNANLKDARLLWPIKOKYGRKISWADLMLTGNVALETMGF 180
Db 112 RGGASTGTQFAPNSNPDNANLKDARLLWPIKOKYGRKISWADLMLTGNVALETMGF 171
Qy 181 KTFGAGGADYWPEDYVWGAETEWLGRKRELENGPLGAYQMGVLIYVPEGPNG 240
Db 172 KTFGGRGVVWPEEDYVWSEKWLASERYSGDRELENGPLAAYQMGVLIYVPEGPNG 231
Qy 241 KPDPKAAARDIRETFRMGNDDEETVALIAGGHTFGKAHGAP-THVGPEPEAPTEAQ 290
Db 232 KPDPKAAARDIRETFRMGNDDEETVALIAGGHTFGKAHGAP-THVGPEPEAPTEAQ 290
Qy 301 SLGKNTYGTGHGADTTTSGLEGAWTKPTQWNSNFFENLFGYEWELTKSPAGAYONKPK 360
Db 291 GLGWISSYKKGKSDTTTSGLEGAWTKPTQWNTSYFDMFLGYDWMWLTSPAGAWQNAV 350
Qy 361 DGAGAGTIPDAHPSKSHAPFMTLTDLALRMDPDYEKISRYYENPDEPDAFAKAWYKL 420
Db 351 DPDEKDLAPDAEDPSKKVPTMMWTTDLALRFDPEYKIRRFHQNPPEFAFAFAWFKL 410
Qy 421 THRDMPKRYLGPVEPOEDLIWQDPIPDVSHPLVDENDIEGLKAKLTESGLTVSELVST 480
Db 411 THRDMPKRYLGPVEPKDFIWDPIPEVDYELT-EAEIEIKAKLTNSGLTVSELVST 469
Qy 481 AWASASTFRNSDRKGGANGARIRLAPOKDWEVNNPOOLARVLTLEGIQEDFNOAQSDNK 540
Db 470 AWASASTFRNSDRKGGANGARIRLAPOKDWEVNEPERLAKVLSVYEDIQREL-----PK 523
Qy 541 AVSLADLIVLAGCAGVEKAAKADAGHEVOVPFNCRADATAEQTQDVEAFEALEPAADGFRN 600
Db 524 KVSADIIVLGGSAAVEKAARDAGFVKVFPFGRGDQEQTDVSEFAVLEPFDGFRN 583
Qy 601 YIRPEKVSABEMLVDRAQLLSAPENTALVGMVRLGTYNGDSHGCVFTNKPGLSND 660
Db 584 YRQEYSVPPPELLVDKQALLGLTAPENTVLGGLRVLGANYRDLPHGVFTDRIGVLTND 643
Qy 661 FFVNLLDLNTKWRASDESDFEGRDPKTGEVKWSGTRVDLIFGNSSELRALAEVYGCAD 720
Db 644 FFVNLLDMNYEWVPTDSG--IYEIRDRKTGEVWTRATRVLDLIFGNSILRSYAEFYAQQD 701
Qy 721 SEEFVKDFVKAQKAVMDLDRFDL 744
Db 702 NQEFVRDFINAWVKVNNADRFDL 725

RESULT 9

Q955N2 ID Q955N2 PRELIMINARY; PRF; 744 AA.
AC Q955N2; DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Peroxidase/catalase (EC 1.11.1.6) (Catalase-peroxidase).
GN CAT.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RA Matsuura T., Miyai K., Trakulnaleaamsai S., Yomo T., Shima Y.,
RA Miki S., Yamamoto K., Urabe I.;
RT "Evolutionary molecular engineering by random elongation
mutagenesis";
RL Nat. Biotechnol. 17:58-61(1998).

CC	-I- CATALYTIC ACTIVITY: 2 H ₂ O(2) = O(2) + 2 H ₂ O.
CC	-I- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL
CC	PEROXIDASE/CATALASE SUBFAMILY.
DR	EMBL: AB020092; BAA37004.1; -
DR	HSSP: P00431; 1BJ9
DR	InterPro: IPR000763; Bac_ctase/prase.
DR	InterPro: IPR002016; Peroxidase.
DR	Pfam: PF00141; peroxidase; 1.
DR	PRINTS: PR00458; PEROXIDASE.
DR	TIGRFAMs: TIGR00198; cat_per_HPI; 1.
DR	PROSITE: PS00435; PEROXIDASE.1; 1.
DR	PROSITE: PS00436; PEROXIDASE.2; 1.
KW	Heme; Hydrogen peroxide; Iron; Oxidoreductase; Peroxidase.
SO	SEQUENCE 744 AA; 83946 MW; 34632D2B5B109BB CRC64;
Query Match	67.8%; Score 2714.5; DB 2; Length 744;
Best Local Similarity	68.7%; Pred. No. 5.5e-167;
Matches 511; Conservative 75; Mismatches 139; Indels 19; Gaps 6	
Qy	1 MENHKSGSSTYNTNTGCKPCTGGSLKQSGAGGTKNRDWPNNMLNLGILRHSSLSDPN 60
Db	1 MEMQ-----NRQNAQCPEHGSVTNQSS-NRTTKNDWPNQLNLSILHQHDKRTNPH 51
Qy	61 DPEDYAEERFKLIDLAAYKKDLAALMTDSODWPPADYCHGPFIRMAHISAGTYRTGDG 120
Db	52 DEEFNYAEERFKLDYALKEDLRLKLTESODWPPADYCHGPTFIRMAHISAGTYRTGDG 111
Qy	121 RGGGSGSQRFAPLNSWPDNANLDKARLLWPDKYGRKISWADLMILGNVALETMGF 180
Db	112 RGNASTGTQRFAPLNSWPDNANLDKARLLWPKKKYGNKISWADLEILAGNVAIESNGG 171
Qy	181 KTFGAGGRADVWEPEEDVYNGAFTWLGDKRYEGDRELENP LGAVOMGLIYVNPGE 240
Db	172 KTIGFGGRVDVWHPEEDVYNGSEKELASERYSGDRELENP LAAVOMGLIYVNPGE 231
Qy	241 KPDPIAARDIRETFEGWMNDEETVALIAGGHTFGKTHCAADAERKYYGPREPAAAGLEEM 300
Db	232 KPDPAARDIRETFRRGMNDEETVALIAGGHTFGKAHGAGPA-TIVGPEPEAPAEIAEQ 290
Qy	301 SLGKNYTGTHGADTTISGLEAKWTKPTQWSNFFENLFGYEWELTKSPAGAYQWKPK 360
Db	291 GLGWISSYKGKGSDDTITSGIEGAWTPPTQWDTSYFDMLFGYDMLTKSPAGAWQMAV 350
Qy	361 DCAGAGTIPAHDPKSKSHAPFLMTTDLALRMDPDYKISRRYENPDEPDADAFKAWYKL 420
Db	351 DPDEKDLAPADEEDSKVPVTMMTDLALRFDPEYKFIARFHQPEFEAEAFARAWFKL 410
Qy	421 THRDMGPKRYLGPVPOEDLWQDPTDYPVSHPLVDENDLEGLKAKLTESGLTWSVSLVST 480
Db	411 THRDMGPKRYLGPVPEKDFIWQDPIPEVDYELT-EAEIEEKAKLTNSGLTVSELVKT 469
Qy	481 AWASASTFRNSDKRGKGANGARILAPQKDEVNPNPOOLARVLTLEIGQDFENQAQSDNK 540
Db	470 AWASASTFRNSDKRGKGANGARILAPQKDEVNEPERLAKVLSYVEDIQREL-----PK 523
Qy	541 AVSLADLIVLAGCAGVEKAAKDAGHEVQVFPNPGRADATAEQTDVEAFEALEPAADGFRN 600
Db	524 KVSITADLIVLGSSAAFEAKAARDGFDYKVPFPFPGRGDQTEQTDVESFAVLEPFDGFRN 583
Qy	601 YIKPEHKVSAEEMLVDRQAQLLSAPETAALVGMRMVLTNGDSQHCVETNPKPQLSND 660
Db	584 YQKQEYSVPPBEELVDRQAQLGLTAPETVLVGLRVLGANGYRDLPHGVFTDRIGVLITND 643
Qy	661 FFVNLDLNTKWRASDESCKVFEGRPDKTGEVQKWSGTRVDLIFGNSSELRALAEVYGCAD 720
Db	644 FFVNLDMNVEWVPTDSG--IYEIRDRKTCGEVRWTATRVDLIFGNSILASAEFYAQDD 701
Qy	721 SEEFVDFDKAKAKVMDLDRFDL 744
Db	702 NQEFVRDFINAWYKVNADRFDL 725

RESULT 10

Q955K7	PRELIMINARY;	PRT;	744 AA.
ID	Q955K7		
AC	Q955K7		
DT	01-MAY-2000 (TremBrel. 13, Created)		
DT	01-MAY-2000 (TremBrel. 13, Last sequence update)		
DT	01-MAR-2003 (TremBrel. 23, Last annotation update)		
DE	Peroxidase/catalase (EC 1.11.1.6) (Catalase-peroxidase).		
GN	CAT.		
OS	Bacillus stearothermophilus.		
OC	Bacteria; Firmicutes; Bacillales; Geobacillus.		
OX	NCBI_TaxID=1422;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Matsuura T., Miyai K., Trakulina I., Yamamoto K., Urabe I.,		
RA	Miki S., Yamamoto K., Urabe I.,		
RT	"Evolutionary molecular engineering by random elongation		
RT	mutagenesis";		
RL	Nat. biotechnol. 17:58-61(1998).		
CC	-/- CATALYTIC ACTIVITY: 2 H ₂ O(2) - O(2) + 2 H ₂ O.		
CC	-/- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL		
CC	PEROXIDASE/CATALASE SUBFAMILY.		
DR	EMBL; AB020067; BAA36979.1; -;		
DR	HSSP; P00431; 1BJ9		
DR	InterPro; IPR000763; Bac_ctase/prase.		
DR	InterPro; IPR002016; Peroxidase.		
DR	Pfam; PF00141; peroxidase; 1.		
DR	PRINTS; PR00458; PEROXIDASE.		
DR	TIGRFAMs; TIGR00198; cat_per_HPI. 1.		
DR	PROSITE; PS00435; PEROXIDASE_1; 1.		
DR	PROSITE; PS00436; PEROXIDASE_2; 1.		
KW	Heme; Hydrogen peroxide; iron; Oxidoreductase; Peroxidase.		
EQ	SEQUENCE 744 AA; 84000 MW; 35CED2142D4F6AB CRC64;		
Query Match 67.8%; Score 2714.5; DB 2; Length 744;			
Best Local Similarity 68.7%; Pred. No. 5.5e-167;			
Matches 511; Conservative 75; Mismatches 139; Indels 19; Gaps 6;			
Qy	1	MENKHSSTYNTGCKPFTGGSLKQSGAGGCTKRDWPNMLNGLIRQHSSLSDPN	60
Db	1	MENO-----NRQNAACQPFHGSVTVNQSS-NFTTKDWWPNOLNLSILHQHDKTNPH	51
Qy	61	DPDPAEAEKFKLLDLAAVKDOLALMTSDSDWPDYGHGPPFIRMAHWSAGTYRIGDG	120
Db	52	DEEFYAEAEQKLDYALKEDRLKLTESQDWWPDYGHGGLFIRMAHWSAGTYRIGDG	111
Qy	121	RGGGGSGRFAPLNSPNDNANLKDARLLWPIKQYGRKISWADLMILCNVALETWGF	180
Db	112	RGGAAGTQRFAPLNSPNDNANLKDARLLWPIKQYGRKISWADLMILAGNVAIESMG	171
Qy	181	KTFGAGGRADVWPEEDYVWGAEHWLGDYKRYGDELENPLGAVOMGLIYVPEGPNG	240
Db	172	KTIFFGGGRVDVWPEEDYVWGSEKWLASERYSGDRELENPLAAGVOMGLIYVPEGPD	231
Qy	241	KPDPIAARADIRETFGRMNDDETVALIAGGHFTGKTGAADAEEKYVVGREPAAGIEM	300
Db	232	KPDPIAARADIRETFGRMNDDETVALIAGGHFTGKAAGAPGPA-TVHVGPEEAPAEQ	290
Qy	301	SLGWKNYTGCHGADTTISGLEAWTKTPTQWSNFFENLFGYEWELTKSPAGAYQWKPK	360
Db	291	GLGWISSYGVKGSDDTTISGLEAWTKTPTQWDTSYFDMLFGYDWMLTKSPAGAWQWAV	350
Qy	361	DGACAGTIPDAHDPKSHAPMLTDLALMDPDYKISRYRYENPDEFADAFKAYKL	420
Db	351	DPDEKADPAEDPSKVPYTMMDTDLALRPDPEYKIAIRFHQNEEFAPAFARWFKL	410
Qy	421	THRDWGPVKRYLGPVPEQEDLIWQDPIPDVSHPLVDENDIEGLKAKILESGLTYSSELVST	480
Db	411	THRDWGPVKRYLGPVPEQEDLIWQDPIPDVSHPLVDENDIEGLKAKILESGLTYSSELVST	469
Qy	481	AWASASTFRNSDKRGANGARIRLAPOKDMEVNNPQOLARVLTKELEGIEDFNQAOQSDNK	540
Db	470	AWASASTFRNSDKRGANGARIRLAPOKDMEVNNPQOLARVLTKELEGIEDFNQAOQSDNK	523

```

QY 541 AVSLADLIVLAGCAGVEKAAKADGAGHEVQVPPNPGRADATBQTDVEAPEALEPADGFRN 600
DB 524 KVSADLIVLGGSAARVEKAAADAGDVKKVPPFPGRGDQTOBQTDVESFAVLEPADGFRN 583
QY 601 YIKPEHKVSAEMLVDRAQLLSAPENTALVGGMRVLGTNYDGSQGVFTNKPGQLSND 660
DB 584 YQKEYSVPPPELLVDRAQLLGLTAPENTVLVGGRLVGLGANYRDLPHGVFTDRIGVLIND 643
QY 661 FFVNLDLNTKWRASDSKVFEGRDKTGVKWSGTRVDLIFGNSSELRALAEVYGCAD 720
DB 644 FFVNLDMNVEWPTDSG--IYEIRDRKTGEVRTATRVLDLIFGNSILRSYAEFYAQDD 701
QY 721 SEEKFKVDFKAAKAVMDLDRFDL 744
DB 702 NQEFVRDFINAWKVMNADRFDL 725

RESULT 11
Q9S5Q5 PRELIMINARY; PRT; 745 AA.
AC Q9S5Q5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Peroxidase/catalase (EC 1.11.1.6) (Catalase-peroxidase).
GN CAT.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RA Matsuura T., Miyai K., Trakulinaeamsai S., Yomo T., Shima Y.,
RA Miki S., Yamamoto K., Urabe I.;
RT "Evolutionary molecular engineering by random elongation
RT mutagenesis."
RL Nat. Biotechnol. 17:58-61(1998).
CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) - O(2) + 2 H(2)O.
CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL
CC PEROXIDASE/CATALASE SUBFAMILY.
DR EMBL; AB020115; BAA37027.1; -.
DR HSSP; P00431; 1bJ9.
DR InterPro; IPR000763; Bac_ctase/prase.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00141; peroxidase; 1.
DR PRINTS; PR00458; PEROXIDASE.
DR TIGRfams; TIGR00198; cat_per_HPI; 1.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS00436; PEROXIDASE_2; 1.
DR Heme; Hydrogen peroxide; Iron; Oxidoreductase; Peroxidase.
SQ SEQUENCE 745 AA; 84135 MW; 9266ED88814B5272 CRC64;

Query Match 67.8%; Score 2714.5; DB 2; Length 745;
Best Local Similarity 68.7%; Pred. No. 5.5e-167;
Matches 511; Conservative 75; Mismatches 139; Indels 19; Gaps 6;

QY 1 MENHKGSGSTYNTNGKCPPTGGSLKQAGGTTKNDWPNMNLGILRHSSLSDPN 60
DB 1 MENQ-----NRQNAQCPFGSVTNQSS-NRTTNKDWPNOLNLSILHQHDKTNPH 51
QY 61 DPFDYAEFKLIDLAARKDLAALMTDSQDMWPADYGHYPFFIRMAHWSAGTYRIGDG 120
DB 52 DEEFNAYAEFKLDYALKEDLRKLTESQDMWPADYGHYPFFIRMAHWSAGTYRIGDG 111
QY 121 RGGGGSGSRFAPLNSWPONANLDKARLLWPITKQYGRKISWADLMILTGNVALETMGF 180
DB 112 RGGASTGTFAPLNSWPONANLDKARLLWPIKKYKGNKISWADLFILAGNVAIESMG 171
QY 181 KTFGAGGRADYWEPEEDYVWGAETEWLGDKRYEGDRELENPLGAVQMGLIYVNPDPNG 240
DB 172 KTIQFGGGRVDVWHPPEDYVWGSKEWLASERYSGDRELENPLAAVQMGLIYVNPDPG 231
QY 241 KPDPFAAARDIRETFGRMAAMDEETVALIAGHTFGRAHGAGPA-THVGPPEAPAIQAD 290
DB 301 SLGKNYTYGTGHGADTTITSGLEGAWTKTPTQWNSNFFENFGYEWELTKSPAGAYQWPK 360
QY 291 GLGWISSYKKGSDTTITSGLEGAWTKTPTQWNTSYDMLFGYDMLTKSPAGAWQWAV 350
DB 361 DGAGAGTIPDAHDPSKSHAPMLTDLALRMDPDYKISRRYENPDEFADAFKAMYKL 420
DB 351 DPDEKDLAPDAEDPSKVPYTMNTTDLALRDPDEVEKIARRFHQNPPEFAFAFAWFKL 410
QY 421 THRDWGPVRYLGPVEPQEDLIWQDPIPDVSHPLVDENDIEGLKAKILESGLTVELSVST 480
DB 411 THRDWGPVRYLGPVEPKEDFIWQDPIPEVDYELT-EAEIEIKAKILNSGLTVSELVKT 469
QY 481 AWASASTFRNSDKRGAGANGARIRLAPQKDMVNNPQOLARVLKLTGEGTQEDFNQAQSDNK 540
DB 470 AWASASTFRNSDKRGAGANGARIRLAPQKDMVNEPERLAKVLSVYEDIQREL-----PK 523
QY 541 AVSLADLIVLAGCAGVEKAAKADGAGHEVQVPPNPGRADATBQTDVEAPEALEPADGFRN 600
DB 524 KVSADLIVLGGSAARVEKAAADAGDVKKVPPFPGRGDQTOBQTDVESFAVLEPADGFRN 583
QY 601 YIKPEHKVSAEMLVDRAQLLSAPENTALVGGMRVLGTNYDGSQGVFTNKPGQLSND 660
DB 584 YQKEYSVPPPELLVDRAQLLGLTAPENTVLVGGRLVGLGANYRDLPHGVFTDRIGVLIND 643
QY 661 FFVNLDLNTKWRASDSKVFEGRDKTGVKWSGTRVDLIFGNSSELRALAEVYGCAD 720
DB 644 FFVNLDMNVEWPTDSG--IYEIRDRKTGEVRTATRVLDLIFGNSILRSYAEFYAQDD 701
QY 721 SEEKFKVDFKAAKAVMDLDRFDL 744
DB 702 NQEFVRDFINAWKVMNADRFDL 725

Query Match 67.8%; Score 2714.5; DB 2; Length 746;
Best Local Similarity 68.7%; Pred. No. 5.5e-167;
Matches 511; Conservative 75; Mismatches 139; Indels 19; Gaps 6;

```


QY 1 MENHKSSSTNTYNTGKCPPTGSLQKOSAGGGTKNRDWPMLNLGILRQHSSLSDPN 60
 DB 1 MENQ-----NRQAAQCPFHGSVTNQSS-NRTTNKDWPNQLNLSILHQHDKRTNPH 51
 QY 61 DPDFYAEFEKKDLAAVKKDLAALMTDSQDWNPADYGHYGPFFIRMAHWSAGTYRIGDG 120
 DB 52 DEEFNYAEFQKLDYALKEDLRLKMTESQDWNPADYGHYGLFIRMAHWSAGTYRIGDG 111
 QY 121 RGGGSGSQRFAPLNSWPDNANLDKARLLWPKIKYGRKISWADLMILTGNALETMGF 180
 DB 112 RGGASTGTQRFAPLNSWPDNANLDKARLLWPKIKYGRKISWADLMILTGNALETMGF 171
 QY 161 KTFEPAGRADWVEEDYWGAEFTWGLDGRYEGDRELEPLGAVQVGLIYVNEPGNG 240
 DB 172 KTIGGGRVDMHPEEDYWGSEKELASERYSGDRELEPLAAVQVGLIYVNEPGDG 231
 QY 241 KPDPTAAARDIRETFGRMAMDEETVALIAGHTFGKTHGAADAKEYVYGRPAAAGIEEM 300
 DB 232 KPDPTAAARDIRETFGRMAMDEETVALIAGHTFGKTHGAADAKEYVYGRPAAAGIEEM 290
 QY 301 SLGWKNTYGTGAGADTITSGLEGAWTKPTQNSNFFENLFGYEWELTKSPAGAYQWPK 360
 DB 291 GLGWISSYKKGKSDTITSGIEGAWTPTPTQWDTSYFDMFLGYDWMWLTSPAGAWQWMAV 350
 QY 361 DGAGAGTTPDAHDPSKSHAPFMTTDLALRMDPDYKISRYYENPDEPDAFAFAKAYKL 420
 DB 351 DPDEKDLAPDAEDPSKVPYTMNTTDLALRMDPDYKISRYYENPDEPDAFAFAKAYKL 410
 QY 421 THRDMPKRVYLGPEVPEQEDLIWQDPIPDVSHPLVDENDIEGLKAKILESGLTVSELVST 480
 DB 411 THRDMPKRVYLGPEVPEQEDLIWQDPIPDVSHPLVDENDIEGLKAKILESGLTVSELVST 469
 QY 481 AWASASTFRNSDKRGANGARIRLAPQKDWENNNPQOLARVLTLEGIQEDFNQAQSDNK 540
 DB 470 AWASASTFRNSDKRGANGARIRLAPQKDWENNNPQOLARVLTLEGIQEDFNQAQSDNK 523
 QY 541 AVSLADLIVLAGCAGVEKAAKADAGHEVQVFPNPGRADATAEOTDVEAFEALEPAADGFRN 600
 DB 524 KVSIAADLIVLGSAAVEKAARDAGFDVKVPFPFGRGDQEQTDVSEFAVLEPFAADGFRN 583
 QY 601 YIKPHEKYSAEMLVDRAQLLSAPMTALVGGMRVLTNYDQSGHQVFTKPKQOLSND 660
 DB 584 YKQEYSVPPPELLYDKAQLGLTAPMTALVGGMRVLTNYDQSGHQVFTKPKQOLSND 643
 QY 661 FVNLILDNTKWRASDESQVFEGRDFTKGVKWSGTRVLDLIFGNSSELRAALAEVYGCAD 720
 DB 644 FVNLILDNTKWRASDESQVFEGRDFTKGVKWSGTRVLDLIFGNSSELRAALAEVYGCAD 701

RESULT 13

Q9S5L8 PRELIMINARY: PRT: 748 AA.
 AC Q9S5L8;
 DT 01-MAY-2000 (TREMELrel. 13, Created)
 DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
 DT 01-MAY-2003 (TREMELrel. 23, Last annotation update)
 DE Peroxidase/catalase (EC 1.1.1.6) (Catalase-peroxidase).
 GN CAT.
 OS Bacillus stearothermophilus.
 OC Bacteria; Firmicutes; Bacillales; Geobacillus.
 OX NCBI_TaxID=1422;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Matsura T., Miyai K., Trakulnaleamsai S., Yomo T., Shima Y.,
 RA Miki S., Yamamoto K., Urabe I.;
 RT "Evolutionary molecular engineering by random elongation
 RT mutagenesis";
 RL Nat. Biotechnol. 17:58-61(1998).
 CC -I- CATALYTIC ACTIVITY: 2 H(2)O(2) - O(2) + 2 H(2)O.

CC -I- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL
 CC PEROXIDASE/CATALASE SUBFAMILY.
 DR EMBL; AB020078; BAA36990.1; -;
 DR HSSP; P00431; 1BJ9.
 DR InterPro; IPR000763; Bac_ctase/prase.
 DR InterPro; IPR002016; Peroxidase.
 DR Pfam; PF00141; peroxidase; 1.
 DR PRINTS; PR00458; PEROXIDASE.
 DR TIGRFAMs; TIGR00198; cat_per_HPI; 1.
 DR PROSITE; PS00435; PEROXIDASE_1; 1.
 DR PROSITE; PS00436; PEROXIDASE_2; 1.
 KW Heme; Hydrogen peroxide; Iron; Oxidoreductase; Peroxidase.
 SQ SEQUENCE 748 AA; 84466 MW; C4BC206B26ED9CC3 CRC64;

Query Match 67.8%; Score 2714.5; DB 2; Length 748;
 Best Local Similarity 68.7%; Pred. No. 5.6e-167;
 Matches 511; Conservative 75; Mismatches 139; Indels 19; Gaps 6;

QY 1 MENHKSSSTNTYNTGKCPPTGSLQKOSAGGGTKNRDWPMLNLGILRQHSSLSDPN 60
 DB 1 MENQ-----NRQAAQCPFHGSVTNQSS-NRTTNKDWPNQLNLSILHQHDKRTNPH 51
 QY 61 DPDFYAEFEKKDLAAVKKDLAALMTDSQDWNPADYGHYGPFFIRMAHWSAGTYRIGDG 120
 DB 52 DEEFNYAEFQKLDYALKEDLRLKMTESQDWNPADYGHYGLFIRMAHWSAGTYRIGDG 111
 QY 121 RGGGSGSQRFAPLNSWPDNANLDKARLLWPKIKYGRKISWADLMILTGNALETMGF 180
 DB 112 RGGASTGTQRFAPLNSWPDNANLDKARLLWPKIKYGRKISWADLMILTGNALETMGF 171
 QY 161 KTFEPAGRADWVEEDYWGAEFTWGLDGRYEGDRELEPLGAVQVGLIYVNEPGNG 240
 DB 172 KTIGGGRVDMHPEEDYWGSEKELASERYSGDRELEPLAAVQVGLIYVNEPGDG 231
 QY 241 KPDPTAAARDIRETFGRMAMDEETVALIAGHTFGKTHGAADAKEYVYGRPAAAGIEEM 300
 DB 232 KPDPTAAARDIRETFGRMAMDEETVALIAGHTFGKTHGAADAKEYVYGRPAAAGIEEM 290
 QY 301 SLGWKNTYGTGAGADTITSGLEGAWTKPTQNSNFFENLFGYEWELTKSPAGAYQWPK 360
 DB 291 GLGWISSYKKGKSDTITSGIEGAWTPTPTQWDTSYFDMFLGYDWMWLTSPAGAWQWMAV 350
 QY 361 DGAGAGTTPDAHDPSKSHAPFMTTDLALRMDPDYKISRYYENPDEPDAFAFAKAYKL 420
 DB 351 DPDEKDLAPDAEDPSKVPYTMNTTDLALRMDPDYKISRYYENPDEPDAFAFAKAYKL 410
 QY 421 THRDMPKRVYLGPEVPEQEDLIWQDPIPDVSHPLVDENDIEGLKAKILESGLTVSELVST 480
 DB 411 THRDMPKRVYLGPEVPEQEDLIWQDPIPDVSHPLVDENDIEGLKAKILESGLTVSELVST 469
 QY 481 AWASASTFRNSDKRGANGARIRLAPQKDWENNNPQOLARVLTLEGIQEDFNQAQSDNK 540
 DB 470 AWASASTFRNSDKRGANGARIRLAPQKDWENNNPQOLARVLTLEGIQEDFNQAQSDNK 523
 QY 541 AVSLADLIVLAGCAGVEKAAKADAGHEVQVFPNPGRADATAEOTDVEAFEALEPAADGFRN 600
 DB 524 KVSIAADLIVLGSAAVEKAARDAGFDVKVPFPFGRGDQEQTDVSEFAVLEPFAADGFRN 583
 QY 601 YIKPHEKYSAEMLVDRAQLLSAPMTALVGGMRVLTNYDQSGHQVFTKPKQOLSND 660
 DB 584 YKQEYSVPPPELLYDKAQLGLTAPMTALVGGMRVLTNYDQSGHQVFTKPKQOLSND 643
 QY 661 FVNLILDNTKWRASDESQVFEGRDFTKGVKWSGTRVLDLIFGNSSELRAALAEVYGCAD 720
 DB 644 FVNLILDNTKWRASDESQVFEGRDFTKGVKWSGTRVLDLIFGNSSELRAALAEVYGCAD 701
 QY 721 SEEFKVFQVFAKAWKMDLDRFDL 744
 DB 702 NQEKVFVRDFINAWKVMNADRFDL 725

RESULT 14
 Q9S5P6

```

ID Q95SP6 PRELIMINARY; PRT; 749 AA.
AC Q95SP6;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Peroxidase/catalase (EC 1.11.1.6) (Catalase-peroxidase).
GN CAT.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RA Matsuura T., Miyai K., Trakunaleaamsai S., Yomo T., Shima Y.,
RA Miki S., Yamamoto K., Urabe I.;
RT "Evolutionary molecular engineering by random elongation
RT mutagenesis.";
RL Nat. Biotechnol. 17:58-61(1998).
CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) - O(2) + 2 H(2)O.
CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL
CC PEROXIDASE/CATALASE SUBFAMILY.
DR EMBL; AB020106; BAA37018.1; -.
DR HSSP; P00431; 1B39.
DR InterPro; IPR00763; Bac_ctase/prase.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00141; peroxidase; 1.
DR PRINTS; PR00458; PEROXIDASE.
DR TIGRFAMs; TIGR00198; cat_per_LPI; 1.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS00436; PEROXIDASE_2; 1.
KW Heme; Hydrogen peroxide; Iron; Oxidoreductase; Peroxidase.
SQ SEQUENCE 749 AA; 84551 MW; 43E089541EA70F19 CRC64;

Query Match 67.8%; Score 2714.5; DB 2; Length 749;
Best Local Similarity 68.7%; Pred. No. 5.6e-167;
Matches 511; Conservative 75; Mismatches 139; Indels 19; Gaps 6;

QY 1 MENKHSGSYNTNTGKCPFTGGSLKQSGAGGTTKRDWPNMLNGLRQHSLSLSDPN 60
DB 1 MENQ-----NRQNAACQPFHGSVTNQSS-NRTTNKDMWPNQLNLSILHODRKTNP 51
QY 61 DPDPDYAEERKDLAALKVKKDLAALMTDSQDWWADYCHYGFPIRMAHWSAGTYRGDG 120
DB 1 DEEFYAEERKLDYALKEDLRKLTMTESQDWWADYCHYGFPIRMAHWSAGTYRGDG 111
QY 121 RGGGSGSQRFAPLNSPNDANLDKARLLWPIKQYGRKISWADLMILTCNVALETWGF 180
DB 112 RGGASTGTQRFAPLNSPNDANLDKARLLWPIKQYGRKISWADLMILTCNVALETWGF 171
QY 181 KTFGAGRADYWEPEEDVYNGAETEWLGDKRYEGDRELENPLGAVQMGIIYVPEGPN 240
DB 172 KTFGAGRADYWEPEEDVYNGAETEWLGDKRYEGDRELENPLGAVQMGIIYVPEGPN 231
QY 241 KPDPIAARDIRETFGRWAMNDEETVALIAGGHTFGKTHGAADAETKVGREPAAGIEEM 300
DB 232 KPDPIAARDIRETFGRWAMNDEETVALIAGGHTFGKTHGAADAETKVGREPAAGIEEM 290

Query Match 67.8%; Score 2714.5; DB 2; Length 750;
Best Local Similarity 68.7%; Pred. No. 5.6e-167;
Matches 511; Conservative 75; Mismatches 139; Indels 19; Gaps 6;

QY 1 MENKHSGSYNTNTGKCPFTGGSLKQSGAGGTTKRDWPNMLNGLRQHSLSLSDPN 60
DB 1 MENQ-----NRQNAACQPFHGSVTNQSS-NRTTNKDMWPNQLNLSILHODRKTNP 51
QY 61 DPDPDYAEERKDLAALKVKKDLAALMTDSQDWWADYCHYGFPIRMAHWSAGTYRGDG 120
DB 52 DEEFYAEERKLDYALKEDLRKLTMTESQDWWADYCHYGFPIRMAHWSAGTYRGDG 111
QY 121 RGGGSGSQRFAPLNSPNDANLDKARLLWPIKQYGRKISWADLMILTCNVALETWGF 180
DB 112 RGGASTGTQRFAPLNSPNDANLDKARLLWPIKQYGRKISWADLMILTCNVALETWGF 171
QY 181 KTFGAGRADYWEPEEDVYNGAETEWLGDKRYEGDRELENPLGAVQMGIIYVPEGPN 240
DB 172 KTFGAGRADYWEPEEDVYNGAETEWLGDKRYEGDRELENPLGAVQMGIIYVPEGPN 231
QY 241 KPDPIAARDIRETFGRWAMNDEETVALIAGGHTFGKTHGAADAETKVGREPAAGIEEM 300
DB 232 KPDPIAARDIRETFGRWAMNDEETVALIAGGHTFGKTHGAADAETKVGREPAAGIEEM 290
```


THIS PAGE BLANK (USPTO)